CLONING AND CHARACTERIZATION OF ZEBRAFISH THYROTROPH EMBRYONIC FACTOR (*tefα* AND *tefβ*)

BY

XINGZHI XU (Bachelor of Medicine)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

NATIONAL UNIVERSITY OF SINGAPORE

1997

DEDICATED TO THE MEMORY OF MY BELOVED

FATHER WHOSE LOVE AND GUIDANCE I WILL

ALWAYS CHERISH.

ACKNOWLEDGMENTS

A special "thank you" is indebted to Dr. Ruowen GE. Thank you for making it possible for me to pursue and finish my M.Sc. degree at your laboratory in the School of Biological Sciences; thank you for your valuable advice and guidance during last two years; thank you for giving me freedom to try new stuff; thank you for your patience and caring. I hope that I have shown you the growth and development as a junior scientist that you anticipated when you accepted me into your lab.

I would like to extend my thanks to Dr. Min WAN, my adviser when I was a special research student in the biochemistry department from Nov. 1994 to Feb. 1995. It is you who gave me the start-point training in the area of molecular biology. I appreciate your training methology. Jacinta, thanks for your help during that period.

I thank Dr. Zhiyuan GONG for providing the adult zebrafish cDNA library, for your allowing me to use equipment at your laboratory, for your advice on some particular techniques. I thank Dr. Eric Weinberg (Department of Biology, University of Pennysylvalia) for the 20-28 hpf cDNA library. Thank Dr. JL DING for your allowing me to use the x-ray developer and the valuable discussion.

I would like to extend my thanks to Kelvin WONG for the cloning of the PCR fragment encoding the 5' end of *tefβ*. Thank Kevin, Dawn, Andrea, the hornours students of 1995-1996 year, for all your help. You guys made the bench work not so boring.

I would like to express my thanks to HweeHoon, for all your help, for all the orderings for me; without your help, my project would not have been done. Thank you for enjoying KFC with me. Zhiqing, Dr. Bala, your help made things easy and smooth, thank you.

I would like to thank Dr. Lucia Pirisi and Dr. Kim Creek for making it possible for me to finish up writing this thesis at your laboratories, School of Medicine, University of South Carolina.

So many thanks are indebted to my wife, Ji LIAO. Thank you for your love and caring and excellent technical supports. Without those, I would not have been here.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Summary

Zebrafish has been shown to be an excellent model to study veterbrate embryonic development. Thyrotroph embryonic factor (TEF) is a member of the PAR (proline- and acid- rich) subfamily of the bZIP (basic leucine zipper) transcription factors. PAR proteins have been shown to have important biological functions such as early embryonic development, liver gene expression, leukemia and apoptosis. We have isolated two cDNA clones (2.7 kb and 1.4 kb) encoding the zebrafish homologues of TEF, designated as *tefα* and *tefβ* respectively. The two *tef*s have distinct amino-terminus and carboxyl-terminus. Compared to the rat TEF gene, their deduced amino acid sequences are highly conserved in the basic DNA binding domain (92% identity), leucine zipper domain (72.1% identity), the PAR region (75% identity), and the transactivation domain (71% identity). Sequence analysis suggests that alternative splicing of putative last exon results in the distinct C-terminus. Genomic Southern analysis indicates that *tef* is a single copy gene in the zebrafish genome, and the two isoforms may use alternative promoters. Northern blot analysis revealed that there is a 2.8 kb transcript corresponding to *tefα* isoform present at abundant levels in adult zebrafish RNA. This message can first be detected at prim-22 stage (35 hours post fertilization). However, no transcript corresponding to *tefβ* can be detected by this assay. RT-PCR analysis can detect both transcripts. Since rat TEF is expressed specifically in the developing anterior pituitary thyrotroph cells during embryogenesis and is thought to play an important role in thyrotropic cell differentiation, the conserved *tef* may also be involved in zebrafish embryonic development. The two

isoforms may have different tissue-specific function and target preferences. The potential function of TEF in programmed cell death is also discussed.

1. Introduction

1.1 Transcription factors

Mammalian genomes are characteristic of huge complexity. Human genome (haploid) has a size of 3 billion base pairs, encoding about 100,000 genes (Lewin, 1994); while zebrafish has a size of 1.7 billion base pairs (Weinberg, 1992). All cell types with a few exception contain the same whole genetic information (Calkhoven and Geert, 1996). Sheep cloning from a somatic cell of a six-year-old ewe provides solid ground that each cell has the potential to express all the genes (Wilmut et al, 1997). Yet each cell type expresses only a unique subset of the total number of available genes (around 10, 000 genes expressed in one cell) (Calkhoven and Geert, 1996). Some genes are strongly expressed, some are always turned on (those are designated as housekeeping genes encoding cellular structural proteins such as actin), some are completely quiescent. Some are expressed in specific tissue(s) and / or at a particular time point during embryonic development. The important generalization is that gene expression is tightly regulated and the phenotype of a cell is determined by the particular gene products it expresses.

Regulation of gene expression can be at transcriptional level generating heteronuclear RNA (hnRNA); at RNA processing level generating messenger RNA; at translational level producing proteins. At each step there can be a balance between synthesis and degradation

that affects the hnRNA pool, the mRNA pool, and finally, the protein pool (Marks et al, 1996). The most important regulation point is transcription (Calkhoven and Geert, 1996).

Eukaryotic transcription is a process of huge complexity. Three RNA polymerases are involved. RNA polymerase I (RNA pol I) transcribes genes encoding the large RNA precursor (45S) that gives rise to 18S, 5.8S, and 28S ribosomal RNA (rRNA) by endonuclease processing. RNA polymerase II (RNA pol II) is responsible for copying genes for proteins, thus it makes mRNA; also it copies a number of genes for small nuclear RNAs (snRNAs). RNA polymerase III (RNA pol III) transcribes only genes for several small RNA molecules, particularly 5S rRNA and tRNA (Marks et al, 1996). RNA pol II is subjected to the most complex control since it transcribes the protein-coding genes (Calkhoven and Geert, 1996). It is well-known that the RNA pol II can not bind to gene promoter region and activate transcription by itself. It can function only in conjunction with a multitude of other proteins (Calkhoven and Geert, 1996; Ranish and Hahn, 1996; Tansey and Herr, 1997). Some of these proteins make up the general transcriptional machinery, which recruits RNA pol II to promoter region and then maintains basal level of transcription. They are designated as basal transcription factors (Ranish and Hahn, 1996; Tansey and Herr, 1997). Some other proteins are required to select the genes to be transcribed at any given moment or in a given cell type and regulate the rate of their transcription. They bind to the *cis* elements or gene enhancer/repressor region which has some distance from the promoter. These proteins are designated as gene-specific transcription factors (Calkhoven and Geert, 1996). Due to the flexibility of DNA

structure, gene-specific transcription factors can interact with the general transcriptional machinery and activate or repress transcription by formation of a loop of flanking DNA region (Lewin, 1994). However, sometimes, a third class of protein factors are required to act as links between the two types of transcription factors , transmitting the gene-specific messages to the general transcriptional machinery. These factors are called coactivators (Kwok, et al, 1994; Kamei et al, 1996).

Transcription factor is a DNA binding protein which can recognize specific DNA sequence at the *cis* elements of a gene (Pabo, 1992). Typically it contains two major domains: DNA binding domain and transactivation domain. Normally DNA binding domain serves to tether the transcription factor to a site on DNA. It often contains helical structures in the core DNA binding region and a cluster of basic amino acids (Pabo, 1992). Transactivation domain is responsible for activating transcription. It often has the structural feature such as acid-rich, proline-rich, glutamine-rich, or serine- and threonine-rich amino acid sequence (He and Rosenfeld, 1991; Pabo, 1992).

Basal transcription factors are factors which bind to TATA box in the TATA containing promoter or initiator binding region in the TATA less promoter and facilitate in recruiting the RNA pol II to maintain low invariant level of transcription. Those includes TFII A, TFII B, TFII D, TFII E, TFII F, TFII H, and TFII J. TFII D consists of TBP (TATA box binding protein) and TAFs (TBP associated factors, at least 14 TAFs have been found in Homo Sapiens). (von Holde and Zlatanova, 1996; Tansey and Herr, 1997). TFII D is

critical to the basal transcriptional machinery. TBP/TAFs complex binds to the core promoter (the minimal DNA sequence required for transactivation in the absence of an activator, consisting of either a TATA element, initiator element, or both elements). TFII D recruits the RNA pol II, TFII B and other transcription factors to form an initiation transcriptional complex to maintain low invariant transcription level. TAFs play a key role as a mediator in this procedure (Tansey and Herr, 1997). TAFs acetylize histones, weakening histone-DNA interaction and then give the DNA access to basal factors. Some TAFs form nucleosome-like octamer allowing TFII D to be incorporated into condensed chromosomes, thereby recognizing promoter DNA for activation. Some TAFs can phosphorylate themselves and basal factors to regulate RNA pol II strength. In some cases, TAFs act directly as a coactivator to interact with gene-specific transcription factors for enhanced or repressed transcription (Tansey and Herr, 1997).

Gene-specific transcription factors can bind to *cis* elements in the vicinity of core promoter region and/or enhancer/silencer which has a distance from the promoter to increase/silence transcription by interacting with the basal transcriptional machinery (Pabo, 1992). The conservation of DNA-binding motifs provides the basis for classification of transcription factors, such as leucine zipper proteins, POU domain proteins, zinc-finger proteins (He and Rosenfeld, 1991; Pabo, 1992). Many of them can form and bind to DNA as homodimer with themselves or heterodimer with the same subfamily members to activate or repress gene expression (He and Rosenfeld, 1991; Pabo, 1992). Gene-specific transcription factors play a key role in development and cell differentiation (He and

Rosenfeld, 1991). For example, rat thyrotroph embryonic factor (TEF) transactivates thyroid stimulating hormone β subunit transcription exclusively at the developing anterior pituitary (Drolet et al, 1991).

As mentioned above, coactivator bridges between the gene-specific transcription factor and the basal transcriptional machinery to transmit signal for activation or repression. Sometimes this bridging is mediated by TAFs. Other times it is mediated by some other mediators (Tansey and Herr, 1997). For example, CREB(cAMP responsive element binding protein) binds to CRE elements. Phosphorylated CREB binds to CBP (CREB binding protein) whose C-terminus interact with the basal transcriptional complex, possibly TFII D to activate the transcription (Kwok et al, 1994; Kamei et al, 1996).

In summary general transcription factors bind to promoter to keep basal level of transcription. Gene-specific transcription factors bind to upstream *cis* elements to activate or repress transcription through coactivators transmitting information to the basal transcriptional machinery. Transcription factors, especially gene-specific transcription factors, play key roles in the spatial and temporal regulation of gene expression.

1.2 Basic region leucine zipper (bZIP) transcription factors

Basic leucine zipper transcription factors (bZIP proteins) belong to a well characterized family of transcription factors. It was first discovered as a conserved sequence pattern in

several eukaryotic transcription factors (C/EBP [CCAAT/enhancer binding proteins], cFos, cJun, and GCN4) (Landschulz et al, 1988). The DNA binding domain of these bZIP proteins generally contain 60-81 residues, and two distinct subdomains: a basic region that contacts the DNA, and the leucine zipper region that mediates dimerization (Pabo, 1992).

The basic region (which contains about 30 residues) is rich in arginines and lysines. It also contains residues that are conserved throughout the family or in a particular subfamily. The basic region is primarily responsible for the sequence preferences of the bZIP proteins (Pabo, 1992).

Leucine zipper sequences are characterized by a heptad repeat of leucines (*a, b, c, d, e, f, g*) over a region of 30-40 residues (Ellenberger, 1994). Hydrophobic and apolar residues predominate at position *a*, and leucine (or valine in some cases) occupies position *d*, and polar and charged amino acids dominate the others (Stuhl, 1989; Vinson, 1993; Ellenberger, 1994). When two bZIP proteins are dimerized, *a* and *d* come into close contact in a side-to-side manner, thus forming a parallel, two-stranded coiled-coil structure. Both interface face the interior of the coiled-coil and form an extensive van der Waals surface. The neighboring charged *e* and *g* residues shield the hydrophobic dimer interface from solvent, completing the knobs-in-holes arrangement of helices in a coiled-coil, and sometimes forming intrahelical or interhelical salt bridge that contributes to dimer stability (Ellenberger, 1994).

Capacity of heterodimerization is a distinct characteristic of bZIP proteins. Residues *e* and *g* of the leucine zipper repeats are particularly important in this regard due to their influence on dimer stability described above (Vinson, 1993; Ellenberger, 1994). Heterodimerization of c-Fos and c-Jun is a good example. Charge repulsion between acidic *e* and *g* residues destabilizes the Fos homodimer and thereby shifts the equilibrium in favor of the Fos/Jun heterodimer (Schuermann et al, 1991; O'Shea et al, 1992). Homodimer and heterodimer can regulate expression of a gene at different degrees or even the opposite way to each other, or regulate expression of different genes (Ellenberger, 1994). Hetrodimerization preference provides the basis for subgrouping of bZIP proteins in addition to sequence homology. Prolin- and acidic- rich (PAR) subfamily is based on sequence homology of PAR region and heterodimerization preference with the same subfamily members (Drolet et al, 1991; Iyer et al, 1991; Hunger et al, 1992).

1.3 PAR subfamily of bZIP proteins

The PAR subfamily of bZIP proteins is composed of three members: DBP, originally described as a factor binding to albumin D-element (Mueller et al, 1990; Lavery and Schibler, 1993; Khatib et al, 1994; Fonjallez et al, 1996); HLF (hepatic leukemia factor), which in its translocated form is involved in acute lymphoblastic leukemia (Inaba et al, 1992; Hunger et al, 1992; Inaba et al, 1994; Falvery et al, 1995); and TEF (thyrotroph embryonic factor), a factor involved in thyrotroph differentiation (Drolet et al, 1991; Khatib et al, 1994; hunger et al, 1996), and its ortholog VBP (vitellogenin-II binding

```
                                    PAR Region
1.pBStef1.9
2.tefα(163-207)   RMTPDPINPDEIEVD VNFEPDPTDLVLSSI PGGELFDPRKHRFSE
3.TEFr(124-168)   RETPSPIDPNCVEVD VNFNPDPADLVLSSV PGGELFNPRKHKFAE
4.TEFh(124-168)   RETPSPIDPNCVEVD VNFNPDPADLVLSSV PGGELFNPRKHKFAE
5.VBPc(176-220)   RNTPSPIDPDCVEVE VNFNPDPADLVLSSV PGGELFNPRKHKFTE
6.HLFh(158-202)   RNTPSPIDPDTIQVP VGYEPDPADLALSSI PGQEMFDPRKRKFSE
7.HLFr(158-202)   RNTPSPIDPDTIQVP VGYEPDPADLALSSI PGPEMFDPRKRKFSE
8.DBPh(188-232)   RDTPSPVDPDTVEVL MTFEPDPADLALSSI PGHETFDPRRHRFSE
9.DBPr(188-232)   RDTPSPVDPDTVEVL MTFEPDPADLALSSI PGHETFDPRRHRFSE

                  PAR                          DNA Binding Domanin
1. pBStef1.9(1-32)                    KV FVPEDQKDDKYWQRR KKNNVAAKRSRDARR
2. tefα(208-252)  EELKPQPMIKKAKKV FVPEDQKDDKYWQRR KKNNVAAKRSRDARR
3. TEFr(169-213)  EDLKPQPMIKKAKKV FVPDEQKDEKYWTRR KKNNVAAKRSRDARR
4. TEFh(169-213)  EDLKPQPMIKKAKKV FVPDEQKDEKYWTRR KKNNVAAKRSRDARR
5. VBPc(221-264)  EDLKPQPMIKKAKKV FVPDEQKDEKYWTRR KKNNVAAKRSRDARR
6. HLFh(203-247)  EELKPQPMIKKARKV FIPDDLKDDKYWARR RKNNMAAKRSRDARR
7. HLFr(203-247)  EELKPQPMIKKARKV FIPDDLKDDKYWARR RKNNMAAKRSREARR
8. DBPh(233-277)  EELKPQPIMKKARKI QVPEEQKDEKYWSRR YKNNEAAKRSRDARR
9. DBPr(233-277)  EELKPQPIMKKARKV QVPEEQKDEKYWSRR YKNNEAAKRSRDARR

                                    Leucine Zipper Dmain
1. tef1.9(33-80)  LKENQITVRAAFLER ENSALRQEVAELRKD FGRCKNTVARYEAKY GAL
2. tefα(253-300)  LKENQITVRAAFLER ENSALRQEVAELRKD FGRCKNTVARYEAKY GAL
3. TEFr(214-261)  LKENQITIRAAFLEK ENTALRTEVAELRKE VGKCKTIVSKYETKY GPL
4. TEFh(214-261)  LKENQITIRAAFLEK ENTALRTEVAELRKE VGKCKTIVSKYETKY GPL
5. VBPc(265-313)  LKENQITIRAAFLEK ENTALRTEVAELRKE VGRCKNIVSKYETRY GPL
6. HLFh(248-295)  LKENQIAIRASFLEK ENSALRQEVADLRKE LGKCKNILAKYEARH GPL
7. HLFr(248-295)  LKENQIAIRASFLEK ENSALRQEVADLRKE LGKCKNILAKYEARH GPL
8. DBPh(278-325)  LKENQISVRAAFLEK ENALLRQEVVAVRQE LSHYRAVLSRYQAQH GAL
9. DBPr(278-325)  LKENQISVRAAFLEK ENALLRQEVVAVRQE LSHYRAVLSRYQAQH GTL
```

Fig. 1 The deduced amino acid sequences of the keep DNA binding domain, and leucine zipper domain of zebrafish *tef* α isoform are compared to those of all the other identified members of the PAR subfamily of bZIP proteins, i.e. rat TEF (TEFr), human TEF (TEFh), chicken vitellogenin binding protein α isoform (VBPc), human hepatic leukemia factor (HLFh), rat hepatic leukemia factor α isoform (HLFr), human albumin D-site-binding protein (DBPh), and rat albumin D-site-binding protein (DBPr).The conserved nine prolines are in bold and underlined, the leucine residues in the leucine zipper domain are in bold and italic letters. The partial clone pBS tef 1.9 encoding the DNA binding and leucine zipper domains of *tefα* is also show in here.

protein), involved in the expression of the chicken vitellogenin gene (Iyer et al, 1991; Burch and Davis, 1994). Each PAR protein can form heterodimer with another PAR protein but not with non-PAR bZIP protein, possibly due to the charged residues at *e* and

g position of leucine zipper domain potentially forming eight electrostatic interactions between the dimerized coiled-coil (Falvey et al, 1995).

Members of this subfamily have the common characteristics of bZIP proteins: a highly conserved DNA binding domain encoding basic-amino-acid-rich region; and a leucine zipper domain serving for dimerization (Cereghini, 1996). They all have four hepatid repeats (***ILLL***, Fig. 1). In addition, sequence homology is extended to a proline- and acidic-rich region located immediately upstream of DNA binding domain, exhibiting 9 conserved proline residues in all members from different species (Fig. 1). Function of PAR domain is unknown. It was proposed to be a transactivation domain as well as harboring inhibitory elements (deletion of PAR region does not affect the transacivation capacity on its responsive-element promoter reporter constructs) (Drolet et al, 1991). Fine mapping of deletion mutants of this region relating to transactivating strength has not been performed yet. Transactivation domain has been localized within a 40 amino acid stretch upstream of PAR region, designated as THAD (TEF/HLF activation domain) (Hunger et al, 1996). This region is highly conserved in all identified PAR members (Fig. 2). PAR proteins all

```
VBPα (313)    83 ASLMPPI WDKTIPYDGESFHLE YMDLDEFLLENGIPS S 120
hTEF (303)    76 ASLMPPI WDKTIPYDGESFHLE YMDLDEFLLENGIPA S 113
rTEF (261)    34 ASLMPPI WDKTIPYDGESFHLE YMDLDEFLLENGIPA S  71
hHLF (295)    64 AFLGPTL WDKTLPYDGDTFQLE YMDLEEFLSENGIPP S 101
rHLF (295)    64 AFLGPTL WDKTLPYDGDTFQLE YMDLEEFLSENGIPP S 101
rDBP (325)    98 SLFAPLL WERTFPF-G---DVE YVDLDAFLLEHGLPP S 131
hDBP (325)    98 GLLAPLL WERTLPF-G---DVE YVDLDAFLLEHGLPP S 131
```

Fig. 2 Multiple alignment of transactivation domain of PAR proteins. Within this region, VBP and hTEF and rTEF, hHLF and rHLF, have the identical sequence, rDBP only has three residues different from hDBP which are underlined. TEF (VBP) shares 71% amino acid identity with HLF, 40% identity with rDBP, 42% identity with hDBP; HLF shares 45% identity with rDBP, 50% identity with hDBP.

bind to a DNA consensus, 5' GTTACGTTAAT-3' (hunger et al, 1996). However, each displays different target promoter preferences and tissue specific expression, presumably due to highly variable N-terminus (Hunger et al, 1996). Rat DBP, TEF, and HLF have been reported to display different circadian expression in liver and different target promoter preferences. TEF transactivates albumin promoter more potently than DBP, while DBP is capable of activating transcription efficiently from the cholesterol 7α hydroxylase (C7αH) promoter (Fonjallez et al, 1996). HLFα stimulates transcription from C7αH promoter more efficiently than from albumin promoter, while the converse is true for HLFβ (Falvery et al, 1995). Since C7αH encodes the rate limiting enzyme in the conversion of cholesterol to bile acid, and follows a similar circadian rhythm as DBP, it was suggested that PAR proteins may be involved in circadian cholesterol homeostasis.

Rat DBP (Mueller et al, 1990) is the first member of the PAR proteins. However, it was not included until rat TEF was cloned and sequence homology alignment showed up the highly conserved PAR domain (Drolet et al, 1991). Although DBP mRNA is ubiquitously expressed in rat adult tissues, significant levels of the transcript are detected in the liver (Mueller et al, 1990; Inaba et al, 1994). DBP binds to albumin and C7αH promoters and activates their transcription (Fonjallez et al, 1996). Since both albumin and C7αH accumulate in a similar circadian rhythm in liver, and C7αH transcript is detected in liver only, DBP is thought to play a key role in cholesterol homeostasis in liver.

HLF was originally isolated due to its presence in a chimerical transcript, E2A HLF, resulted from action of the t(17; 19)(q22; p13) chromosomal translocation (Inaba et al, 1992). It drives the leukaemic transformation of early B-cell precursor (Inaba et al, 1992). Recently E2A-HLF has been reported to be an apoptosis suppresser, indicating that PAR proteins may be involved in the cell death process (Inaba et al, 1996). Rat HLF gene encodes two transcriptional activators by alternative usage of promoters (Falvery et al, 1995). Both can activate albumin and C7αH but at different preference levels (Falvery et al, 1995).

The recently cloned C. elegans CES-2 (for CEll-death Specification) encodes a bZIP protein. Its DNA binding domain and leucine zipper domain are most similar to PAR proteins, and it has identical DNA binding sequence (Metzstein et al, 1996). Our analysis further reveals that there is a 36 amino acid stretch right upstream of the basic region which has a 30% identity to PAR region of TEF, and 3 conserved proline residues (Fig. 3). CES-2 is a transcriptional repressor of programmed cell death. The similarity between CES2 and PAR proteins indicates that the bZIP domain of CES-2/PAR protein may be involved in apoptosis.

1.4 Thyrotroph embryonic factor (TEF)

Rat TEF was cloned as a transcription factor that binds to and transactivate TSHβ (Thyroid stimulating hormone, β subunit) promoter (Drolet et al, 1991). Its transcript

```
gi|1543073 (U60979) CES-2 [Caenorhabditis elegans]
           Length = 211

 Score = 177 (79.5 bits), Expect = 4.5e-17, Sum P(2) = 4.5e-17
 Identities = 37/71 (52%), Positives = 49/71 (69%)

Query:   199 PQPMIKKAKKVFVPEDQKDDKYWQRRKKNNVAAKRSRDARRLKENQITVRAAFLERENSA 258
             PQ    +   V +PE++KD  Y++RR+KNN AAKRSRDARR KE QI  +A  LEREN
Sbjct:    98 PQRSPSRKMSVPIPEEKKDSAYFERRRKNNDAAKRSRDARRQKEEQIASKAHALERENMQ 157

Query:   259 LRQEVAELRKD 269
             LR +V+ L ++
Sbjct:   158 LRGKVSSLEQE 168

 Score = 40 (18.0 bits), Expect = 4.5e-17, Sum P(2) = 4.5e-17
 Identities = 11/36 (30%), Positives = 16/36 (44%)

Query:   174 LVLSSIPGGELFDPRKHRFSEEELKPQPMIKKAKKV 209
             L   S+P   L  P K   E+EL   P+  ++  V
Sbjct:    55 LPFDSLPTTNLLTPTKKIKLEDELCASPVSSRSSTV 90
```

Fig. 3 Sequence homology analysis of zebrafish *tef*β (query) and CES2 (subject)by BLAST program. It demonstrates that DNA binding domain and leucine zipper domain (top panel) have 52% identity and 69% positive. Also partial PAR domain of *tef*α (36/50) has 30% identity and 44% positive to CES2 protein. In addition, there are three conserved proline residues within this region (the bold and underlined letters).

appears first exclusively in a small rostral part of developing anterior pituitary gland corresponding to the precise area of thyrotroph development on the day 14 (e14) (Drolet et al, 1991). This restricted pattern of TEF gene expression is maintained till e16. It was thought to be a thyrotroph specific transcription factor during embryogenesis. It could heterodimerize with DBP (Drolet et al, 1991) which was the basis for the establishment of this new subfamily. In adult tissues, it is ubiquitously expressed with relatively high level in the liver, the brain, and the kidneys(Drolet et al, 1991; Fonjallaz et al, 1996). Rat TEF also binds to and transactivates transcription of albumin and C7αH genes (Fonjallez et al, 1996). Expression of rat TEF in a circadian rhythm correlates with DBP to tune liver homeostasis, especially in the metabolism of cholesterol and bile salts (Fonjallez et al, 1996). Human TEF has been cloned by homology screening (Fonjallez et al, 1996). It has the same DNA binding sequence and transcriptional properties as rat TEF (Khatib et al,

1994) and HLF (Hunger et al, 1996). It has been localized to chromosome 22q13 (Khatib et al, 1994). It is widely expressed in adult tissues with relatively high level in the liver and kidneys (Khatib et al, 1994). TEF mRNA was also found at modest levels in several different human ALL (acute lymphoblast leukemia) cell lines and in HepG2 hepatocellular carcinoma cell line (Hunger et al, 1996). It suggests that TEF may be involved in tumorigenesis.

The chicken ortholog of TEF, VBP, was cloned as a factor which binds to vitellogenin gene promoter (Iyer et al, 1991). It has overall 88% amino acid identity to rat TEF (Burch and Davis, 1994). Due to alternative usage of promoter and alternative splicing of last exon, it encodes four isoforms owing to different combinations of the two distinct N-terminus and the two distinct C-terminus. The four isoforms show the same binding affinity in vitro, but have different tissue-specific expression (Burch and Davis, 1994). Alternative usage of promoters may be a common event of PAR proteins since it happens to HLF as well (Falvery et al, 1995). Function of VBP is still quite ambiguous.

1.5 Zebrafish is an excellent vertebrate model in embryonic developmental studies

Systematic mutant analysis has revealed many genes required for invertebrate embryonic development in Drosophila malanogaster (Nusslein-Volhard and Wieschaus, 1980). Molecular characterization of mutants has helped to resolve mechanisms of cell-cell

interactions that lead to lineage development and pattern formation. Though many of the fundamental principles of vertebrate embryology have been elucidated by classical experiments in amphibian and avian, neither Xenopus nor chick is suited to carry out genetic studies to provide insights into the genetic control of early development (Eisen, 1997). Obviously mouse is a good candidate because of 1) pretty well understood early embryology; 2) molecular characterization of some mutants; 3) the availability of the improving lineage genetic map and physical map. However, large-scale mutant screen has been considered to be too slow and expensive, mainly owing to small litter sizes, large genome size, the high cost of housing large numbers of animals, and the inability to observe and manipulate all stages of the developing embryo easily (Russant and Hopkin, 1994).

While people were looking around, in 1981 the work of G. Streisinger drew people's attention to the potential of zebrafish (Danio rerio) as an effective experimental system which is highly suited to genetic analysis (Streisinger et al, 1981). Owing to the recognized overwhelming advantages over mouse in early embryogenesis studies, it has become an excellent model in vertebrate embryogenesis. First genetic linkage map (Postlethwait et al, 1994) is being improved, physical map is on its way, large-scale systematic mutant screen has been helping to understand mechanisms of development.

Zebrafish is a stripped 2-inch-long fresh water tropical fish. Compared to mouse, its size is much smaller, therefore it is possible to house a large number of animals more cheaply; it

also has a small size of genome, around 1.7 billion base pairs (haploid), with a chromosomal number of 25 (haploid). Importantly one female can produce as many as 1,000 eggs at one time mostly after the onset of a light cycle. The eggs are externally fertilized immediately by sperm released into water by a male fish at 28.5°C. Distinctly the haploid embryos progress almost normally through early development, allowing investigators to detect recessive early-acting mutations without breeding them to homozygosity. Most importantly, the transparency of the embryo and speed of its early development and maturation favor genetic studies tremendously (Weinberg, 1992; Westerfield, 1995; Eisen, 1997). The time from fertilization to gastrulation is only about 5 hours at 28.5°C; pairs of somites appear at an interval of approximately 30 minutes between 10 and 24 hours; and by 24 hours postfertilization, a recognizable animal with rudimentary eyes and brain and beating heart is wiggling its tail; it reaches sexual maturity in as little as 2-3 months. It is completely feasible to breed four generations per year (Weinberg, 1992).

Owing to the above mentioned advantages over mouse as a model animal, zebrafish is important for understanding molecular genetic mechanisms underlying vertebrate development. Since PAR proteins are involved in the anterior pituitary development and programmed cell death and tumorigenesis, it is important and interesting to isolate and characterize zebrafish homologues of PAR proteins and study their function in zebrafish development. In this thesis, I have isolated two cDNA clones encoding two isoforms of zebrafish *tef* transcripts, with identical transactivation domain, PAR domain, DNA binding

region, and leucine zipper domain, but distinct N-terminus and C-terminus. Southern blots and sequence analysis revealed that the two isoforms may result from alternative promoter usage, alternative splicing of the last exons and differential polyadenylation. Northern analysis detected *tefα*, a 2.8kb mRNA which first appears at prim-22 stage (35 hours post fertilization). RT-PCR detected both transcripts from total adult RNA. The function and significance of *tefs* in zebrafish are discussed.

2. Materials and Methods

2.1 Plasmid preparation

Small scale preparation of plasmid was made using modified CTAB-DNA precipitation method (Del Sal et al, 1989). This method is based on findings that the cationic detergent cetyl-trimethyl ammonium bromide (CTAB) binds to nucleic acids in the presence of high concentration of sodium salts and the CTAB-nucleic-acid complex releases nucleic acids at low concentration of sodium salts (< 0.6 M).

The 4-ml overnight culture in Luria-Bertani (LB) medium with appropriate antibiotics was harvested in 15-ml falcon tube by centrifugation at 4,800 rpm for 5 minutes using the GS-15R centrifuge(Beckman, USA). The bacterial pellet was resuspended in 400 µl STET buffer (8% sucrose, 50 mM Tris pH 8.0, 50 mM EDTA, 0.1% v/v Triton X-100) and transferred to 1.5 ml microtube. Then 10 µl of 50 mg/ml lysozyme (mucopeptide N-acetylmuraloylhydrolase from chicken egg white, Sigma, USA) solution was added the bacterial cell suspension and mixed well by vortexing. After incubation at room temperature for 5 minutes, the sample was boiled for 45-60 seconds to stop the lysis reaction. The bacterial debris was pelleted by centrifugation at 14,000 rpm for 10 minutes. The floppy pellet was removed by a sterile toothpick, then10 µl of 10 mg/ml RNase A (ribonuclease A from bovine pancreas, Sigma, USA) solution was added to the supernatant, and the tube was incubated at 68°C for 10 minutes. The DNA was

precipitated by the addition of 20 µl of CTAB solution (5% hexa-decyltrimethyammonium bromide w/v). The CTAB-DNA pellet was recovered by centrifugation for 5 minutes at 14,000 rpm. Four hundred µl of 1.2M sodium chloride solution was used to resuspend the CTAB-DNA pellet and DNA itself was re-precipitated using 1 ml of pure ethanol. DNA pellet was washed in 0.5 ml of 70% ethanol followed by centrifugation at 12,000 rpm for 5 minutes. After removing the ethanol, DNA was dried in a SpeedVac for 5 minutes before being dissolved in 20-30 µl of water. All the centrifugation was performed at room temperature.

DNA was quantified by optical density reading taken at 260nm (OD_{260nm}) and 280 nm (OD_{280nm}). One unit of OD_{260nm} is equivalent to 50 µg /ml of double strand DNA, and a OD_{260nm} : OD_{280nm} ratio between 1.7 - 2.0 indicates acceptable quality. Normally 4-5 ml overnight culture can yield 50 - 80 µg of high copy plasmid. DNA now is ready for restriction enzyme digestion or sequencing.

2.2 Genomic DNA preparation

2.2.1 *E. Coli* genomic DNA preparation

Considering that commercial salmon sperm DNA is not a good source as a non-specific blocker in hybridization for zebrafish research, we prepared sonicated DH5α DNA as the blocking agent. In the early morning one fresh DH5α colony was inoculated into 20 ml of LB broth. It was incubated at 37°C with shaking at 250 rpm for around 8 hours. Ten ml of

the culture was reinoculated to 1 liter of LB broth and incubated overnight at 37°C with shaking at 250 rpm. In the following morning, cells were pelleted at 7000 rpm for 10 minutes (JA 14 rotor, Beckman, USA). The pellet was resuspended in 50 ml of cold STE buffer (100 mM NaCL, 10 mM Tris, pH 8.0, 1 mM EDTA, pH 8.0). After being repelleted at 7000 rpm for another 10 minutes (JA 20 rotor, Beckman, USA), the pellet was resuspended in 50 ml of TGE buffer (50 ml buffer contains 0.49 gram of Dextrose, 1.25 ml of 1M Tris-HCl, pH 8.0, and 2.0 ml of EDTA, pH 8.0), then 625 µl of 100 mg/ml lysozyme was added, and the mixture was incubated at 37°C for half an hour. After that, 10% SDS was added to a final concentration of 0.5%. The suspension was mixed well before proteinase K was added to a final concentration of 0.1 mg/ml. Incubation at 65°C for 1 hour was followed. The sample was sheared through 20 Gauge needle, and then centrifuged at the speed of 7,000 rpm for 10 minutes. The supernatant was extracted by Phenol:Chloroform:Isoamyl alcohol (25:24:1) twice and Chloroform:Isoamyl alcohol (24:1) once. DNA was precipitated down by addition of pure ethanol, rinsed by 70% ethanol, and dissolved in appropriate volume of TE buffer. The genomic DNA was sonicated into fragments less than 3 kb examined by agarose gel electrophoresis. The sonicated DNA was kept at 4°C or -20°C for future use. Typically 1 litre overnight culture could yield about 10 mg DNA.

2.2.2 Zebrafish genomic DNA isolation

Mammalian DNA is usually extracted by digestion of tissues or cells with proteinase K in the presence of EDTA and SDS, followed by extractions with phenol. This method yields

DNA whose size (100-150 kb) is adequate for Southern analysis (Sambrook et al, 1989). Basically we followed the protocols in the Molecular Cloning (Sambrook et al, 1989). One piece of adult zebrafish (around 200 mg) was sacrificed in the liquid nitrogen and grounded into powder. Two ml of extraction buffer (10 mM Tris-Cl, pH8.0, 0.1 M EDTA, pH 8.0, 20 μg /ml pancreatic RNase, and 0.5% SDS) was added and mixed well with the tissue powder. Proteinase K was added to a final concentration of 100 μg /ml. Then incubation at 50°C for 3 hours with periodically swirling was followed. This viscous solution was extracted three times by phenol equilibrated by Tris-Cl (pH 8.0), avoiding too much mechanical shearing. DNA was precipitated by ethanol, using a glass rod to loop out DNA, and rinsed in 70% ethanol. Extracted DNA was dissolved into 1 ml of TE buffer. The DNA concentration was 1.9 μg /μl. It was ready for digestion and Southern analysis.

2.3 Restriction enzyme digestion

Restriction digestion is used to screen recombinant clones, release specific DNA fragments, or to get restriction enzyme map of interested fragments. All of the restriction enzymes used in this project were from the New England Biolabs, USA except *EcoRI* and *XhoI* which were from Promega, USA. All digestions were performed at 37°C for 1-3 hours with proper restriction buffer in the presence of bovine serum albumin (BSA) to inhibit protease activity, except for *BstX I* which is digested at 55°C. Buffer II from the New England Biolab was found to be good for *EcoRI* and *XhoI* activity. Normally 3 units

of restriction enzyme were used to digest 1 μg of plasmid DNA. The samples were then analyzed by electrophoresis using appropriate concentration of agarose gel dependent on the size of the fragments for separation. The interested fragments were recovered by GenClean II kit (Bio101, USA).

2.4 Recovery of DNA fragments from agarose gel

GenClean kit II (Bio 101, USA) was used to recover the interested DNA fragments ranging from 200 bp to 10 kb from agarose gel according to manufacturer's instructions. This kit is based on the facts that its specially formulated silica matrix, GLASSMILK, binds single and double strand DNA without binding DNA contaminants, and the GLASSMILK-DNA complex releases DNA into low salt solution such as TE buffer or water at 45-55°C.

Briefly, the gel piece containing the interested DNA band was melted at 55°C in a chaotropic agent, 6 M sodium iodide. The volume of the sodium iodide was approximately three times of the quantity of the gel piece. Typically five to ten microlitre of GLASSMILG was used for 1 μg or less DNA, the suspension was incubated with the DNA-agarose solution for 5 minutes at room temperature to allow DNA binding to the matrix. The GLASSMILK-DNA complex was pelleted by spinning at 10, 000 rpm for 10 seconds, the pellet was then washed in 700 μl of NEW washing solution twice. After removal of the washing solution as much as possible, 10 μl of water was added to mix

with the GLASSMILK-DNA complex and mixed well. The mixture was then incubated at 55°C for 2-5 minutes. By centrifugation at 10,000 rpm for 30 seconds, the supernatant contained the released DNA fragments for further manipulations.

2.5 Screening of zebrafish cDNA library

2.5.1 Information of the zebrafish cDNA libraries

The adult and 20 - 28 hour post fertilization cDNA libraries were kind gifts of Dr. Zhiyuan GONG (School of Biological Sciences, National University of Singapore) and Dr. Eric Weinsberg (Department of Biology, University of Pennsylvania, USA) respectively. Both were constructed in the Unit - ZAP XR vector system (Stratagene, USA) which allows rapid *in vivo* excision of pBluescript SK(-) phagemid. Double digestion of the phagemid with *EcoRI* and *XhoI* will release the DNA insert.

2.5.2 Titration of cDNA library

2.5.2.1 Preparation of plating bacteria

XL1 - Blue MRF' is recommended by the manufacturer as host strain for the Uni - ZAP bacteriophage lambda. One single colony of XL1 - Blue MRF' was inoculated into 20 ml of LB medium supplemented with 0.2% maltose and incubated overnight at 37°C with moderate agitation. In the following morning the cells were pelleted by centrifugation at 3,000 rpm for 10 minutes. The cell pellet was resuspended in about 10 ml of sterile 10 mM magnesium sulfate. OD_{600nm} measurement helped to adjust the cells to an appropriate

density, normally OD_{600nm} = 2, i.e., 1.6 x 10^9 cells/ml). The cell suspension could be kept at 4°C for up to three weeks.

2.5.2.2 Plating bacteriophage lambda for titration

Titration reveals how many plague forming units (pfu) in 1 ml of bacteriophage (pfu/ml). Ten fold serial dilutions of bacteriophage stocks in SM buffer (0.58% w/v NaCl, 0.2% w/v $MgSO_4.7H_2O$, 50 mM Tris-Cl pH 7.5, 0.01% w/v gelatin) were prepared. One hundred microlitre of each dilution was incubated with the same volume of the plating bacteria at 37°C for 15 minutes to allow the bacteriophage particles to adsorb to the bacterial. After incubation the solution was mixed with 3 - 5 ml of melted (47°C) NZCYM top agar (1% tryptone, 0.5% NaCl; 0.5% bacto-yeast extract, 0.1% casamino acids, 0.2% $MgSO_4.7H_2O$, 0.7% bacto-agar, Difco. USA), and poured into the 82 mm NZCYM agar plate (1.5% bacto-agar, rest are the same as NZCYM top agar). When the top agar was solidified, the plates were inverted and incubated at 37°C for around 7 hours. Counting the number of isolated plagues in the appropriate dilution plate revealed the titer of the cDNA library. The titres for the adult and the embryonic cDNA libraries are 1.5 x 10^9 and 1.5 x 10^9 pfu/ml respectively.

2.5.3 Preparation of membrane lifts for screening

Hybond-N (Amersham) nylon disc membranes were used for screening, 132-mm membranes for primary screening, and 82-mm membranes for secondary screening. For primary screening, thirty microlitres of 10^{-3} dilution of both cDNA libraries equal to 4.5 x

10^4 pfus was grown on the 145-mm NYCYM plate, 10 plates for each library were prepared. As soon as pin-sized plagues could be observed and before they cross-touched, the plates were transferred into 4°C fridge overnight.

For the secondary screening, the positive plagues were picked into microtubes containing 1 ml of SM buffer and incubated at room temperature at least for 2 hours. Normally one plague can release 10^7 bacteriophages. Ten microlitres of 10^{-3} dilution were grown on 85-mm NZCYM plate, which was around 100 bacteriophages. When the isolated plagues appeared whose diameter was around 0.5 mm, the plates were transferred into fridge overnight.

In the following morning duplicate membrane lifts were prepared for each plate. The first lift was kept touched with plagues for two minutes, the second one for five minutes. DNA-side-up membranes were left on Whatman 3MM filter paper for air dry. DNA was denatured in denaturation solution (1.5 M NaCl and 0.5 M NaOH) for 5 minutes, neutralized in neutralization buffer (1.5 M NaCl and 0.5 M Tris-HCl pH 7.5) for 5 minutes. Membranes were then equilibrated in 2 x SSC solution (for 20 x SSC solution, 3 M NaCl, 0.3 M sodium citrate pH 7.0) for 15 minutes, and dried on Whatman 3MM filter paper for 5 - 10 minutes. DNA was fixed to membrane by UV crosslinking using Spectrolinker XL-1500 UV crosslinker (Spectronics, USA), twice of DNA side and once the other side at the optimal crosslinking condition (1.2×10^5 uJ/cm^2). Membranes now are ready for hybridization.

2.5.4 Probe labeling and purification

2.5.4.1 Random-primed labeling of the probes

Random hexanucleotides have been used as primers in a method for labeling DNA fragment to produce high-specific-activity probes. This procedure relies on the ability of primers to anneal to multiple sites along the length of a DNA template. The primer-template complex represents a substrate for Klenow fragment of DNA polymerase I. The enzyme synthesizes new DNA fragment by incorporating nucleotide monophosphates at the 3'-OH group provided by the primers. The newly synthesized DNA is made radioactive by substituting a radiolabeled nucleotide for a non-radioactive one in the reaction mixture. Prime-It II random primer labeling kit (Stratagene, USA) had been used to label all the probes for cDNA library screening, Southern hybridization, and Northern hybridization. The kit uses a 3' exonuclease-deficient mutant of Klenow fragment of DNA polymerase I and random nanomer instead of hexamer. It yields high specific-activity probes with relatively short time, normally 20 minutes including column purification.

Essentially, the instruction manual of the kit was followed. DNA fragments ranging from 86 bps to 2,000 bps were pulled out from vectors by appropriate restriction enzyme digestion. They were recovered from agarose gel using GeneClean kit. Those were used as templates. Twenty five ng of template was denatured with 10 µl of random 9-mer primers (27 OD_{260nm} unit/ml) in boiling water for 5 minutes. At this point the total volume was topped up to 34 µl by H_2O. Annealing was done by briefly leaving the reaction tube at

room temperature, and 5 µl of 10 x dCTP buffer containing dATP, dGTP and dTTP (0.1 mM of each) was added to the tube and mixed. After brief centrifugation to collect all the solution into the bottom, 5 µl of α-^{32}P dCTP at 3,000 Ci/mmole (Amersham, USA) and 1 µl of Exo(-) Klenow enzyme (5u/µl) were added. The tube was incubated at 37-40 °C in a thermomixer (Eppendorf, Germany) for 15 minutes. The reaction was stopped by adding 2 µl of stop mix (0.5M EDTA, pH 8.0). The mixture now was ready for column purification.

2.5.4.2 Nick column purification

Probe purification to get rid of unincorporated radio-labeled nucleotides will yield high signal-to-background ratios on hybridization results. Nick column (Pharmacia Biotech, USA) prepackaged with Sephadex G-50 (DNA grade) was employed for this purpose. Elutants are determined by the molecular weight of DNA fragments. Single buffer, TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) was used for column equilibration and elution. Briefly, the column was equilibrated by 3 ml of TE buffer, then 50 - 100 µl of sample was loaded, three consecutively 400-µl elutants were collected into three separate microtubes. The second elutant contained most of the radioactivity (50,000 to 150,000 cpm/µl) and was used for hybridization. Normally hybridization buffer contained probe at the concentration of 1.0 x 10^6 cpms per ml.

2.5.5 Hybridization

2.4.5.1 Prehybridization

Non-specific blocking by sonicated *E.Coli* DNA and Denhardt's solution, as well as probe purification, is essential to obtain high signal-to-background ratio. Prehybridization buffer contained 6 x SSC (for 20 x SSC, 3M NaCl, 0.3 M tri-sodium citrate, pH7.0), 5 x Denhardt's reagent (for 50 x Denhardt's solution, 0.2% bovine serum albumin, 0.2% Ficoll400, and 0.2% polyvinylpyrroidone), 0.5% sodium dodecyl sulfate (SDS), and denatured *E.Coli* DNA (150 μg per ml of hybridization buffer). Membranes were held in the hybridization bottle and separated by meshes (Hybaid, UK). Prehybridization, hybridization, and post-hybridization wash were performed inside the oven (hybaid, UK). Temperature was dependent on the aqueous buffer (68°C) or the hybridization buffer containing 50% formamide (42°C). Prehybridization was carried out for 1 to 4 hours.

2.5.5.2 Hybridization

Probe was denatured in the boiling water for 10 minutes and chilled in ice for 2 minutes. Appropriate amount of probe was added directly into prehybridization buffer to obtain radioactivity of 1.0×10^6 cpm per ml of buffer. Hybridization was carried out for 20 hours.

2.5.5.3 Post-hybridization washes

Adequate post-hybridization washes can get rid of non-specifically bound probes; on the other hand extensive washing will remove specific bindings as well. It is important to monitor the membranes through washes by mini Geiger counter. Background at 0 to 2 cps is acceptable, while scattered spots at 5 or more cps indicate positive signals. Stringent wash was performed for all hybridizations. After hybridization unbound probes were

removed by incubating and rolling twice with 1 x SSC, 0.1% SDS for 15 minutes at room temperature. Non-specifically bound probes were got rid of by incubating the membranes twice in 1 x SSC and 0.1% SDS for 20 minutes at 68°C. If background was still high, further wash by 0.1 x SSC and 0.1% SDS at 68°C was necessary with frequently monitoring by Geiger counter.

2.5.5.4 Autoratiography

After washing, excess liquid was removed by blotting the membranes on Whatman 3MM paper. The membrane was placed between two sheets of Saran Wrap in a cassette with intensifying screens. Exposure was undertaken at -80°C for 72 hours. Film developing was performed using the Kodak automatic X-ray developer (Kodak, USA).

2.5.6 *In vivo* excision of pBluescript phagemid using ExAssist/XLOLR system

Uni-ZAP for the construction of the zebrafish cDNA libraries has been designed to allow simple, efficient *in vivo* excision and recirculization of any cloned insert contained within the lambda vector to form a phagemid containing the insert. The ExAssist/XLOLR system (Stratagene, USA) allows efficient excision of pBluescript phagemid from the Uni-ZAP vector, while eliminating problems associated with helper phage co-infection. The ExAssist helper phage contains an amber mutation that prevents replication of the phage genome in a nonsuppressing *E.Coli* strain such as XLOLR cells. This allows only the excised phagemid to replicate in the host, getting rid of the possibility of productive co-infection from the ExAssist helper phage.

Essentially, we followed the manufacturer's instructions. An overnight culture of XL-1 blue MRF' (Stratagene, CA, USA) was grown in LB broth at 37°C. In the following morning, a 1% reinoculation of XL-1 Blue MRF' and an inoculation of a fresh XLOLR colony in LB broth were incubated at 37°C with shaking at the speed of 150 rpm. XL-1 blue cells were grown to mid-log phase (OD_{600nm} around 0.3 and 0.4) after 2.5 hours. The culture was pelleted by centrifugation, and bacteria were resuspended in 1 mM $MgSO_4$ at OD_{600nm} around 1.0. XLOLR cells were allowed to grow for 6 hours to OD_{600nm} around 0.7 to 0.9, and then left at room temperature for transformation usage. One hundred thousand phage particles from a positive plague from secondary screening were incubated with 200 µl of XL-1 blue cells and 1 µl of ExAssist helper phage (10^6 pfus) at 37°C for 20 minutes. Further incubation for 2.5 hours at 37°C with shaking at 250 rpm was followed after addition of 3 ml of LB broth. Heating inactivation of the phage was performed at 70°C for 15 minutes. After centrifugation the supernatant contained excised phagemids which contained inserts. Two hundred µl of freshly grown XLOLR cells were transformed by 10 µl of excised phagemids and incubated at 37°C for 20 minutes. After that, 300 µl of LB broth was added, and the tube was incubated at 37°C at 180 rpm for 1.5 hours. Bacteria (60 µl, 1/10 volume; and 540 µl, 9/10 volume) were plated on two individual LB-Ampicillin (ampicillin concentration is 100 µg /ml) plates. The plates were incubated at 37°C overnight. Dozens of colonies were typically observed on the 1/10 plate. Colonies were now ready for insert-size checking by polymerase chain reaction (PCR) and plasmid preparation.

2.6 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is a powerful tool to amplify DNA fragments defined by one pair of primers millions of times by thermostable DNA polymerase. It was extensively used in this project for colony screening and rapid amplification of cDNA ends (RACE). Information of the primers used was provided in Table 1.

2.6.1 Recombinant-colony screening

Theoretically, PCR is sensitive enough to detect a single DNA molecule; and denaturation at high temperature will release plasmids from cells. One pair of primers whose sequences are on the vector flanking cloned insert will define the size of the insert by a PCR reaction. It is much more sensitive, less time-consuming than restriction enzyme digestion analysis of mini-prep plasmids. In brief, all the colonies to be examined were marked sequentially. The target colony was touched by a sterile toothpick. The toothpick then scratched the bottom of a PCR tube. Twenty μl of PCR mix was then added to each tube, overlaid by one drop of mineral oil (Sigma, USA). One hundred μl of PCR mix contained 1.5 units of Taq DNA polymerase, 10 μl of 10 x PCR buffer, 6 μl of 25 mM $MgCl_2$, 4 μl of 2.5 mM dNTP mix (all solutions are from Promega, USA), and 50 pmole of each reverse and forward primers. For PCR-Script vector and pBluescript vector (Stratagene, USA), T3 and T7 were used as sense and anti-sense primers respectively; for pGEM-T vector (Promega, USA), T7 and SP6 were used as forward and reverse primers respectively.

Table 1 Primer information

Reference name	(A/S)*	Sequence (5′→3′)	Sequence location	Applications
TEFA1	(A)	CTG ATT CTC TTT CAG GCG	927 to 944 of *tefα*	5′RACE of *tefβ* and RT-PCR
TEFA2	(A)	CCA GTA TTT ATC ATC CTT CTG	855 to 875 of *tefα*	5′RACE of *tefβ* and RT-PCR
TEFA5′UTR	(S)	CGG CTT TTT GTG CAG CGA GTA AC	134 TO 156 of *tefα*	RT-PCR of *tefα*
TEFSEQS	(S)	CAG CGG AGG AAG AAG AAC	735 to 752 of *tefβ*	sequencing of *tefβ*
TEFSEQA	(A)	GCC GTT GCA TTA TAC G	1113 to 1128 of *tefβ*	sequencing of *tefβ*
TSK1	(S)	GAT TAA CCG TCT CCA GCA CAC TGC	11 to 34 of *tefβ*	3′RACE of *tefβ* transcript
TSK2	(S)	CGG GTG TTG ATC AAC TGA AAC CG	43 to 65 of *tefβ*	RT-PCR of *tefβ* transcript
TEFSX	(S)	ATG AAG CCT ATT TCC ATC ACG	173 to 193 of *tefα*	RT-PCR of *tefα* transcript
ZACTIN2	(S)	GAT GAG TCT GGA CCA TCC ATC GTC	ACTINA and ZACTIN2 flanks 560 bp fragment of zebrafish β actin 3′ UTR for RT-PCR control	
ACTINA	(A)	GAA TTA GTC AGT GTA CAG G		

A/S: antisense or sense primer

PCR program included an initiate denaturation at 94°C for 5 minutes, followed by 30 cycles each of which was programmed as denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and elongation at 72°C for 1 minute. PCR was performed on PE480 themocycler (Perkin Elmer, USA). Fifteen μl of PCR products was analyzed by 1.0 to 1.5% agarose gel electrophoresis. Colonies which yielded expected size of PCR products were inoculated for plasmid preparation.

2.6.2 Rapid amplification of cDNA ends (RACE)

The cDNA libraries used in this projects were constructed in the Uni-ZAP vector (Stratagene, USA). The cDNAs were cloned uni-dirctionally into *EcoRI* site and *XhoI* site (5'→3') of polylinker of pBluescript SK(-), which was incorporated into Uni-ZAP vector. These libraries were good templates for both 5'RACE and 3'RACE. The combination of antisense gene-specific primer and T3 or SK primers is suitable for 5'RACE, and combination of sense gene-specific primer and T7 or M13-20 primers fits 3'RACE. RACE PCR reaction (50 μl each) contained 2.5 units of Taq DNA polymerase, 5 μl of PCR 10 x buffer, 3 μl of 25 mM $MgCl_2$, 4 μl o 2.5 mM dNTP, 50 pmole each of one pair of appropriate primers. PCR programs was dependent on the gene-specific primer's melting temperature (Tm). Twenty μl of PCR product was electrophoresized on 1.0 to 1.2% agarose gel. Significant bands were recovered by the Genclean kit and used for cloning.

2.7 Cloning

2.7.1 Subcloning

In this project, three types of cloning were employed, subcloning, PCR cloning, and nest deletion cloning. Subcloning was done typically for generation of plasmid containing shorter or specific internal fragment of insert for DNA analysis and probe generation. If vector and cloning fragment were generated by the same single restriction enzyme, dephosphorylation of vector was necessary to reduce significantly self ligation of vector. Dephosphorylation was performed at 37°C for 30 minutes in a 20 µl reaction which contained 2 µl of shrimp alkaline phosphatase buffer (10 x, 200 mM Tris-HCl, pH 8.0, 100 mM $MgCl_2$) and 1 unit of shrimp alkaline phosphatase (Amersham, USA). The reaction was terminated by incubation at 70°C for 10 minutes. Enzyme digested open vector and the same enzyme cut fragment were ligated by T4 DNA ligase (Promega, USA). Ligation reaction was typically 10 µl, containing 3 units of DNA ligase, 1 µl of 10x ligation buffer (10 x, 300 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP). The molar ratio of insert and open vector was varied from 1-3:1. Ligation was incubated at 14°C for at least 3 hours, for both sticky-end and blunt-end ligation.

2.7.2 PCR cloning

Target fragments recovered by Gen-cleaned kit from PCR products were directly cloned into the pGEM-T vector system(Promega, USA). The pGEM-T vector is prepared by cutting Promega's pGEM-5Zf(-) vector with *EcoRV* and adding a 3' terminal thymidine to both ends. These single 3'-T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmid because generally PCR generates 3' adenine

overhang products. In addition, the vector contains T7 primer at the 5' end and SP6 primer at the 3' end of insert site. It is suitable for further analysis of recombinant DNA such as colony PCR screen and sequencing. Ligation reaction and condition were almost the same as those of subcloning, with the exception of insert and vector molar ratio from 8:1 to 1:8 dependent on the quantity of PCR product.

2.7.3 Nest deletion cloning

Due to the limit of manual sequencing by sequenase version 2.0 T7 DNA polymerase (Amersham, USA), and the limit of availability of appropriate restriction enzymes within insert for subcloning, it is desirable to construct plasmids containing progressive unidirectional deletions of interested insert. The Erase-a-Base system (Promega, USA) was used for this purpose to successfully sequence *tef* clones. The system is based on the facts that exonuclease III (*Exo III*) can specially digest DNA from a 5' protruding or blunt end at an uniform rate at a preset temperature, and that mung bean nuclease can cleave single strand DNA after *Exo III* treatment. Religation will yield plasmid containing shorter insert. Combination of time and temperature control for ExoIII treatment will generate any desired length inserts convenient for sequencing.

We strictly followed the technical manual provided by the manufacturer. In order to sequence *tefα* in pBluescript SK(-) vector, *BstX I* was used to generate ExoIII resistant 3' overhang at the 5' end of *tefα*, *BamHI* was used to produce 5' overhang accessible by ExoIII. Twenty μg of pBS *tefα* was incubated with 40 unit of *BamHI* and 40 units of

BstX I (New England Biolabs, USA) at 37°C for 1.5 hours (optimal temperature for *BamHI*) and followed by incubation at 55°C for 1.5 hours (optimal temperature for *BstX I*). Agarose gel electrophoresis of 1 µl of digestion reaction demonstrated complete digestion. The digest was extracted by phenol:chloroform:isoamyl alcohol(25:24:1) once and by chloroform:isoamyl alcohol (24:1) once, precipitated down by addition of 0.1 volume of 3 M NaOAc and 2 volumes of pure ethanol. After the sample was incubated at -20°C for half an hour, the sample was centrifuged at 14,000 rpm at 4°C for 15 minutes. The pellet was washed by 70% ethanol, and dissolved into 20 µl of H_2O. The deletion mix contained 5 µl of DNA, 6 µl of the 10 x ExoIII buffer (10 x, 660 mM Tris-HCl, pH 8.0, 6.6 mM $MgCl_2$), and 49 µl of water. The tubes were warmed in the thermomixer at 30°C, then 500 units of ExoIII were added. Twenty time points with 40-second interval (theoretically 140 bp deletion per time point) were set. Two and half µl at each point was transferred into prechilled tube containing 7.5 µl of S1 nuclease mix (for 25 time points, 172 µl of water, 60 units of S1 nuclease, 27 µl of S1 7.4 x buffer (7.4 x, 0.3M KOAc, pH 4.6, 2.5 M NaCl, 10 mM $ZnSO_4$, 50% glycerol). After all samples were collected, S1 tubes were incubated at room temperature for 30 minutes. S1 nuclease was inactivated by addition of 1 µl of stop buffer (0.3M Tris base, 0.05 M EDTA). Three µl of each sample was analyzed on a 1% agarose gel. Based on the electrophoresis information, 11 time points (1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20) were selected for Klenow polishing. One µl of Klenow mix(for 20 µl mix, 19 µl of Klenow buffer[1 x, 20 mM Tris-HCl, pH 8.0, 100 mM $MgCl_2$], 3 units of Klenow DNA polymerase) was added to each sample. The tubes were incubated at 37°C for 3 minutes, and then 1 µl of dNTP mix (0.125 mM of each) was

added and incubated for additional 5 minutes at 37°C. The Klenow was heat-inactivated at 65°C for 10 minutes. The samples now were ready for ligation. Forty µl of ligation mix (for 1 ml, 100 µl of ligation 10 x buffer, 100 µl of 50% PEG, 10 µl of 100 mM DTT, 790 µl of water, 5 units of T4 DNA ligase) was added. The tubes were incubated at room temperature for one hour. The samples were ready for transformation.

2.7.4 Transformation

2.7.4.1 Preparation of competent cells

Efficiency of transformation is critical for success of cloning. Normally 10^6 to 10^7 transformed colonies per microgram of supercoiled plasmid is good enough for most cloning except library construction. Competent cells prepared using calcium chloride is a simple method, and its transformed efficiency is around 5×10^6 to 2×10^7 (Sambrook, 1989). In this project, we prepared DH5α and XL-1 blue competent cells by this method with minor modifications, and got the same efficiency. Briefly, single fresh colony was inoculated into 3 ml of LB broth and incubated at 37°C with shaking at 250 rpm overnight. In the following morning, 0.5 ml of the overnight culture was re-inoculated into a 250 ml flask containing 50 ml of LB broth with continuous shaking at 250 rpm at 37°C, the culture was monitored by OD_{600nm} reading until it reached between 0.4 and 0.7. Cells were pelleted by centrifugation at 2500 rpm (swing rotor) at 4°C for 7 minutes. The pellet was resuspended in 20 ml of ice-cold 100 mM $CaCl_2$. The suspension was repelleted at the same condition. The pellet was resuspended in 20 ml of $CaCl_2$, and kept in ice for 30 minutes. Again bacteria were pelleted by centrifugation at 2500 rpm at 4°C for 5 minutes.

The pellet was resuspended into 1.6 ml of 100 mM $CaCl_2$ and 0.4 ml of pure glycerol. One hundred µl was aliquoted into a prechilled tube, and then transferred into -80°C deep freezer. All the procedures had to be performed at 4°C to get better transformation efficiency. One hundred µl of competent cells was good for one transformation.

2.7.4.2 Transformation

Five µl of prechilled ligation reaction was mixed with 100 µl competent-cell tube in a prechilled microtube. The tube was kept on ice for 30 minutes, then heat-shocked at 42°C for 50 seconds, and chilled on ice for 2 minutes. Half ml of prewarmed LB broth (42°C) was added. Incubation at 37°C with 180 rpm was performed to recover Ampicillin resistant gene carried by the plasmids. Transformants were screened by double selection with antibiotics and white/blue selections. Transformants were plated on the LB/Amp(100 µg /ml)/IPTG(0.5 mM)/X-Gal(80 µg /ml) plates. The plates were incubated at 37°C overnight. White colonies were picked for PCR colony screening.

2.8 DNA sequencing analysis

2.8.1 Sequencing reaction

The enzymatic dideoxynucleotide chain termination protocol (Sanger et al, 1979) was used for double stranded plasmid sequencing. It is based on the facts that DNA polymerases must have a primer to begin synthesis; and that incorporation of dideoxynucleotide will terminate further elongation because dideoxynucleotide does not

have 3' OH for elongation; and that resolution in a denaturing acrylamide/bisacrylamide gel can differentiate one base pair difference. We used sequenase version 2.0 DNA sequencing kit (Amersham, USA) for sequencing. Since we used CTAB prepared plasmid for sequencing instead of single strand DNA, modified sequencing protocol (Del Sal et al, 1989) was followed to fit CTAB-prep plasmids. Briefly, 5 μg of plasmid and 5 pmole of sequencing primer were mixed in a microtube and denatured in 0.1 N NaOH (total volume was 10 μl) at 68°C for 10 minutes. The denaturation was neutralized by adding 4 μl of TDMN buffer (0.28 M TES-free acid, 0.12 M HCl, 0.05 M DTT, 0.08 M $MgCl_2$, and 0.2 M NaCl), and then incubated at room temperature for 10 minutes. Labeling reaction was done by adding 5.5 μl of labeling mix (3 μl of diluted labeling dGTP mix (dilution factor was 5), 2 μl of diluted sequenase (dilution factor was 8 in the sequenase dilution buffer, and 0.5 μl of α-S^{35}-dATP (10 μCi/μl, Amersham, USA), and by incubating at room temperature for 5 minutes. Then 3.5 μl of labeling reaction was transferred to each prewarmed (37°C) 2.5 μl termination mix (ddATP, ddGTP, ddCTP, or ddTTP). The mixture was incubated at 37°C for 5 minutes for chain termination. The termination reaction was stopped by adding 4 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). The reactions now were ready for electrophoresis.

2.8.2 Sequencing gel electrophoresis

We used 6% denaturing acrylamide/bisacrylamide gel containing 6% of acrylamide/bisacrylamide (29:1), 8 M of urea. For 60 ml of gel solution, 42 μl of TEMED

(Life Technology, USA) and 420 µl of 10% ammonium persulfate were used for polymerization. The thickness of gel was 0.35 mm or 0.4 mm depending on the sequencing apparatus used. The sequencing reactions for one sample were loaded in the order of ddA, ddG, ddC, and ddT next to one another. Electrophoresis was performed at the fixed power of 50 to 60 watts to maintain gel temperature below 50°C. The length of electrophoresis time was dependent on the sequencing data we needed. Normally a short run was performed for 2.5 hours and a long-run for 5 hours, which could provide sequencing information up to 200 and 350 bps respectively. This sequence length was the up-limit of the sequencing kit. Once electrophoresis was done, gel was transferred to Whatman 3MM filter paper, and dried at 80°C for 1 hour (Hoffer, USA). The dried gel was exposed to Kodak X-ray film (Kodak, USA) at room temperature overnight.

2.8.3 Sequence analysis

Nucleotide sequence and its deduced amino acid sequence were analyzed by BLAST(http://dot.imgen.bcm.tmc.edu:9331/seq-search/Options/blast.html) or FASTA(http://www.ebi.ac.uk/searches/fasta.html) programs. The latter allows gaps in the sequence, while the former does not allow gaps. Sequence comparison was performed by BCM multiple sequence alignments (http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html).

2.9 Isolation of total RNA

Total RNA was used throughout this project. Single step method isolation by acid guanidinium thiocyanate-phenol-chloroform extraction has been well developed (Chomczynski and Sacchi, 1987; Chomczynski, 1993). TRIzol reagent (Life Technology, USA) we used includes phenol and guanidine thiocyanate in a mono-phase solution. Tissue or egg is homogenized in the reagent and the homogenate is separated into the aqueous and organic phases by addition of chloroform and centrifugation. RNA remains exclusively in the upper aqueous phase and is precipitated by addition of isopropanol.

In detail, 200 mg of zebrafish tissues or 100 eggs were quickly frozen in the liquid nitrogen and homogenized in 1 ml of TRIzol reagent. The homogenate was transferred to a microtube and incubated at room temperature for 5 minutes to allow nucleoproteins to dissociate. Then 200 µl of chloroform was added, the mixture was vortexed vigorously for 2 minutes. The tube was incubated at room temperature for another 3 minutes, and then centrifuged at 12, 000 g for 15 minutes at 4°C to separate aqueous and organic phases. Five hundred µl of aqueous phase was transferred to a new tube, and equal volume of isopropanol was added and mixed well. RNA was pelleted by centrifugation at 12, 000 g for 10 minutes at 4°C. The pellet was washed in 1 ml of 70% ethanol, and air dried. The pellet was dissolved into 25 to 30 µl of RNAase-free water. RNA was quantified by OD_{260nm} reading (1 OD_{260nm} is equivalent to 40 µg/ml of RNA), and qualified by OD_{260nm} / OD_{280nm} ratio at around 1.7 to 1.9. Further examination of RNA integrity was required to check degradation on a formaldehyde agarose gel.

2.10 Reverse transcription PCR (RT-PCR)

Manipulation of nucleic acids are done easier with DNA than RNA because most organisms use DNA as the genome, and restriction enzymes operate on DNA and not RNA. However, in mammalian cells, mRNA is most related to the protein, also it is not as complex as the chromosomal DNA which contains more than 90% non-transcribed sequence (Lewin, 1994). Reverse transcriptase, derived from retrovirus which transcribes its RNA genome into DNA and integrates into host genome, make test-tube synthesis of DNA from RNA possible. RT-PCR has been widely used to detect specific transcript and for gene cloning. In this project, we employed RT-PCR for examination of *tef* transcripts.

Access RT-PCR system (Promega, USA) was used in this project. It is designed for the reverse transcription and PCR amplification of specific target transcript from total RNA. This system uses AMV reverse transcriptase from Avian Myeloblastosis Virus for the first strand DNA synthesis, and the thermostable Tfl DNA polymerase from Thermus flavus for second strand cDNA synthesis and DNA amplification. By using an optimized single-buffer system, this method permits extremely sensitive detection of RNA transcripts. In addition, the improved performance of AMV reverse transcriptase at elevated temperature (48°C) minimizes problems encountered with secondary structures in RNA. In detail, in a 50 µl reaction in a 0.5 ml microtube, 1 µg of total RNA was used as template, mixed with 10 µl of 5 x AMV/Tfl reaction buffer, 1 µl of dNTP mix (10 mM each dNTP), 2 µl of 25 mM $MgSO_4$, 50 pmole of each downstream and upstream primers, 1 µl of AMV reverse

transcriptase (5 units per μl), 1 μl of Tfl DNA polymerase (5 units per μl). The final volume was topped up to 50 μl by adding water. One drop of mineral oil was overlaid on the solution surface. Reverse transcription and PCR amplification were done on the PE 480 themocycler (Perkin Elmer, USA). The program we used was reverse transcription at 48°C for 60 minutes, AMV RT inactivation and RNA/cDNA/primer denaturation at 94°C for 2 minutes, PCR amplification by 35 cycles each of which included denaturation at 94°C for 30 seconds, annealing at 60°C for 1 minutes, extension at 68°C for 2 minutes, and followed by one cycle of final extension at 68°C for 7 minutes. Then 10 to 15 μl of product was analyzed on agarose gel electrophoresis.

2.11 Northern analysis

Once a gene has been cloned, it is useful to determine the size of the message and quantitate the specific message compared to a housekeeping standard message such as actin. It is helpful to confirm that the cDNA clone contains full length transcript and to predict protein coding region. The method used to analyze RNA in this way is Northern analysis, in which total or poly (A)+ mRNA is run on a denaturing agarose gel and detected by hybridization of a labeled probe on a membrane. Details are described below.

2.11.1 Agarose/formaldehyde gel electrophoresis of RNA

A 1.2% agarose gel in 1 x MOPS (10 x, 200 mM MOPS, 10 mM EDTA, 50 mM NaOAc, pH 7.0) and 6% formaldehyde solution was prepared, gel running buffer containing 1 x

MOPS and 3% formaldehyde was made. Gel preparation and electrophoresis were done inside the chemical hood. RNA sample was denatured at 68°C for 10 minutes by combining 2.5 µl of RNA up to 10 µg and 7.5 µl of formamide, 3 µl of 37% formaldehyde, 1.5 µl of 10 x MOPS, 0.5 µl of 10mg/ml ethidium bromide. RNA ladder (Gibco, USA) was treated at the same way. After denaturation the tubes were chilled on ice for 2 minutes. Then the samples and RNA ladder were loaded each with 1 µl of loading dye. The gel was run at 100 volts (or 8 volts per cm length of gel) till first dye near the end. The gel was rinsed with distilled water twice to remove formaldehyde, and finally soaked in 25 mM PB buffer (0.5 M Na_2HPO_4, 0.57 M NaH_2PO_4). The gel now was ready for RNA transfer to a nylon membrane.

2.11. 2 Transfer of RNA to Hybon-N membrane

Seven hundred ml of 25 mM PB was poured into an appropriate size of plastic container. One layer of Whatman 3 MM paper was put as a wick on a glass. The gel was placed on the paper with bottom side up, and then covered with one identical size of Hybond-N membrane (Amersham, USA). Three layers of prewet Whatman 3 MM paper were laid. All the procedures should not produce any air bubbles between layers of paper and membrane. Finally a pack of paper towels was put on the Whatman 3 MM paper and a glass at the top loaded by a lead O-ring around 1 kilograms. The transfer was performed overnight. After transfer, RNA was fixed to membrane by crosslinking once at the RNA side at the optimal dosage by Spectrolinker XL-1500 UV crosslinker (Spetronics, USA). The membrane now was ready for hybridization. Hybridization procedures are identical to

cDNA library screen described previously. Probes for Northern analysis were 5' untranslated region (5'UTR) specific to *tefα* (1 to 236 th bp of *tefα*), *tefα* 2 kb probe (1 to 2074 th bp of *tefα*, which overlaps partially with *tefβ*), and 3'UTR 0.8 kb probe (1522 to 2358 th bp). Probe labeling was described above.

2.12 Genomic Southern analysis

Southern analysis is applied to localize particular sequence within genomic DNA. In this project we also used the method to show that the two isoforms of *tef* use alternative promoters. Genomic DNA is digested by one or more restriction enzymes and separated on the agarose gel and transferred to a nylon membrane. Radiolabeled DNA fragment is used to locate the complimentary bands in the genomic DNA.

2.12.1 Restriction enzyme digestion and gel electrophoresis of genomic DNA

Fifty μg of zebrafish genomic DNA was digested individually at a volume of 50 μl by *EcoRI* (150 units), *EcoRV* (150 units), *BamHI* (150 units), *HindIII* (150 units), and *PstI* (150 units) at 37°C overnight. Digested DNA was separated on 1.0% agarose gel using TAE buffer at 3 volts per cm length gel for a better separation. When loading dye moved near the end, electrophoresis was stopped. The gel was subjected to denaturation and transfer.

2.12.2 Transfer of DNA to membrane

The gel was soaked in 0.125 N HCl for 20 minutes for depurination to increase transfer efficiency, and then rinsed with water. The gel was then denatured in denaturation solution (1.5 M NaCl and 0.5 M NaOH) for 30 minutes with gentle agitation. After denaturation, the gel was rinsed with water, and then neutralized in neutralization buffer (1.5 M NaCl and 0.5 M Tris-HCl, pH 7.5) for 30 minutes with gentle agitation. After neutralization, the gel was equilibrated in 2 x SSC solution. Transfer procedures and hybridization were identical to Northern analysis. Probes we used were *tefα* 2.0 kb (1 to 2074 th bp of *tefα*) and *tefβ* N-terminal specific fragment (1 to 86 th bp of *tefβ*). Probe labeling was described before.

3. RESULTS AND DISCUSSION

3.1 A 1.9 kb cDNA encodes the DNA binding domain and leucine zipper domain of zebrafish *tef*

During the course of setting up conditions for cDNA library screening using digoxigenin-labeled DNA probes, we picked out one false-positive plague from the 20-28 hpf (hour post fertilization) embryonic cDNA library. This plague was excised into pBluescript SK(-) plasmid. Enzyme digestion by *EcoRI* and *XhoI* generated two fragments whose sizes were around 1.2 kb and 0.7 kb. Digestion by *EcoRI* itself pulled out one fragment of 1.2 kb (Data not shown). It indicated that the size of this insert was around 1.9 kb with an internal *EcoRI* site. Sequencing of 5' end of this 1.9 kb clone by T3 primer revealed 253 bp nucleotides (Fig. 4A) which is highly homologous to rat TEF, human TEF, chicken TEF homologue VBP, human HLF, rat DBP, and human DBP (Fig. 4B). Its deduced amino acid sequences were shown in Fig. 5A. Those sequences indicated that frame 3 was a possible open reading frame (ORF) because it has a continuous stretch of amino acids not interrupted by nonsense codon except at the end of C-terminal, which is probably the stop codon of the full transcript. BLAST search demonstrated similar results of homology. The ORF 3 has 80% identity and 93% similarity with chicken VBP (Fig. 5B). Since all the genes having high homology with this 1.9 kb cDNA belong to proline- and acidic- rich basic leucine zipper transcription factor (PAR bZIP proteins), it suggests that this partial cDNA clone may encode one member of zebrafish PAR bZIP proteins, and probably TEF.

A.

```
agaaagtgtttgtgcctgaggatcagaaggatgataaatactggcagcggaggaagaagaacaatgttgcag
ccaaacgatcacgggatgccaggcgcctgaaagagaatcagatcacagttcgagcggcatttctggagagag
aaaactcagcgctcagacaggaagtggcagagttgcgcaaggactttgggcgctgcaagaacacagtggcac
gatatgaagccaaatacggagcgctgtaaggggcaat (253 bps)
```

B.

Sequences producing	High-scoring Segment Pairs:	High Score	Probability P(N)	N
gb|S58745|S58745	thyrotroph embryonic factor=leucine z...	685	1.8e-49	1
gb|U06935|HSU06935	Human thyrotroph embryonic factor (TE...	667	6.0e-48	1
gb|U44059|HSU44059	Human thyrotroph embryonic factor (TE...	667	6.6e-48	1
gb|U09221|GGU09221	Gallus gallus White Leghorn beta/beta...	632	5.9e-45	1
emb|X68985|HSHEPLF	H.sapiens mRNA for hepatic leukemia f...	572	6.1e-40	1
gb|M95585|HUMHLF	Human hepatic leukemia factor (HLF) m...	572	7.6e-40	1
dbj|D28468|HUMDBPD	Human mRNA for DNA-binding protein TA...	515	3.6e-35	1
gb|U06936|HSU06936	Human albumin D-box binding protein (...	514	4.4e-35	1
gb|U79283|HSU79283	Human albumin D-box binding protein m...	514	4.5e-35	1
gb|M95586|HUME2AHLF	Human E2A/HLA fusion protein (E2A/HLF...	510	1.1e-34	1
gb|J03179|RATDBPA	Rat D-binding protein mRNA, complete ...	456	2.6e-29	1

```
gb|S58745|S58745 thyrotroph embryonic factor=leucine zipper
           transcription factor [rats, pituitary, mRNA, 817 nt]
           Length = 817

  Plus Strand HSPs:

 Score = 685 (189.3 bits), Expect = 1.8e-49, P = 1.8e-49
 Identities = 185/245 (75%), Positives = 185/245 (75%), Strand = Plus / Plus

Query:     1 AGAAAGTGTTTGTGCCTGAGGATCAGAAGGATGATAAATACTGGCAGCGGAGGAAGAAGA 60
             ||||||| ||||| ||||| || |||||||||| || ||||||    || | |||||||
Sbjct:   573 AGAAAGTCTTTGTCCCTGATGAGCAGAAGGATGAGAAGTACTGGACAAGGCGTAAGAAGA 632

Query:    61 ACAATGTTGCAGCCAAACGATCACGGGATGCCAGGCGCCTGAAAGAGAATCAGATCACAG 120
             ||||||| ||||| ||||| || ||||| |||||||||||||| |||||||||||||
Sbjct:   633 ACAATGTGGCAGCTAAACGCTCCCGGGACGCCAGGCGCCTGAAGGAGAATCAGATCACCA 692

Query:   121 TTCGAGCGGCATTTCTGGAGAGAGAAAACTCAGCGCTCAGACAGGAAGTGGCAGAGTTGC 180
             | || || |||||||||||||  || ||| |||| ||  |   ||| || || ||| | |
Sbjct:   693 TCCGGGCAGCATTTCTGGAGAAGGAGAACACAGCCCTGCGGACGGAGGTTGCTGAGCTTC 752

Query:   181 GCAAGGACTTTGGGCGCTGCAAGAACACAGTGGCACGATATGAAGCCAAATACGGAGCGC 240
             |||||||  | ||    ||||||| ||  ||| |    |||||  |||| |||||  |
Sbjct:   753 GCAAGGAGGTGGGCAAGTGCAAGACCATCGTGTCCAAGTATGAGACCAAGTACGGGCCCT 812

Query:   241 TGTAA 245
             |||||
Sbjct:   813 TGTAA 817
```

Fig. 4 Nucleotide sequence of 5' end of the 1.9 kb cDNA clone reveals high homology with rat TEF, human TEF, chick TEF homologue VBP, human HLF, human DBP, and rat DBP, which are members of proline- and acidic rich (PAR) bZIP proteins. A. nucleotide sequence of 5' end (253 bps). B. Nucleotide database search through BLAST program shows 75% identity and 75% positive with rat TEF. Query is the nucleotide sequence of the 5'end of the 1.9 kb cDNA clone, while subject is the rat TEF sequence.

A.

5'3' Frame 1

R K C L C L R I R R Met I N T G S G G R R T Met L Q P N D H G Met P G A **Stop** K R I
R S Q F E R H F W R E K T Q R S D R K W Q S C A R T L G A A R T Q W H D Met K P
N T E R C K G Q

5'3' Frame 2

E S V C A **Stop** G S E G **Stop** **Stop** I L A A E E E E Q C C S Q T I T G C Q A P E R E S D
H S S S G I S G E R K L S A Q T G S G R V A Q G L W A L Q E H S G T I **Stop** S Q I R S
A V R G N

5'3' Frame 3

K V F V P E D Q K D D K Y W Q R R K K N N V A A K R S R D A R R L K E N Q I T V
R A A F L E R E N S A L R Q E V A E L R K D F G R C K N T V A R Y E A K Y G A L
Stop G A

B.

Sequences producing High-scoring Segment Pairs:		High Score	Probability P(N)	N
pir||S50109	vitellogenin binding protein, beta/b...	341	1.1e-42	1
sp|P41224|TEF_RAT	THYROTROPH EMBRYONIC FACTOR /pir||A4...	338	3.2e-42	1
gi|1181892	(S79820) hepatic leukemia factor bet...	333	1.7e-41	1
gi|181912	(M95586) E2A/HLF fusion protein [Hom...	298	1.4e-34	1
pir||A55558	albumin D-box binding protein - huma...	275	9.4e-32	1

```
pir||S50109 vitellogenin binding protein, beta/beta isoform - chicken
            gi|483938 (U09221) vitellogenin binding protein (VBP), beta/beta
            isoform [Gallus gallus]
            Length = 293

 Score = 341 (155.3 bits), Expect = 1.1e-42, P = 1.1e-42
 Identities = 63/78 (80%), Positives = 73/78 (93%)

Query:     1 KVFVPEDQKDDKYWQRRKKNNVAAKRSRDARRLKENQITVRAAFLERENSALRQEVAELR 60
             KVFVP++QKD+KYW RRKKNNVAAKRSRDARRLKENQIT+RAAFLE+EN+ALR EVAELR
Sbjct:   208 KVFVPDEQKDEKYWTRRKKNNVAAKRSRDARRLKENQITIRAAFLEKENTALRTEVAELR 267

Query:    61 KDFGRCKNTVARYEAKYG 78
             K+ GRCKN V++YE +YG
Sbjct:   268 KEVGRCKNIVSKYETRYG 285


sp|P41224|TEF_RAT THYROTROPH EMBRYONIC FACTOR pir||A40579
            trans-activating transcriptional regulatory protein TEF - rat
            gi|237085 (S58745) thyrotroph embryonic factor,  TEF=leucine zipper
            transcription factor [rats, pituitary, Peptide, 261 aa] [Rattus sp.]
            Length = 261

 Score = 338 (153.9 bits), Expect = 3.2e-42, P = 3.2e-42
 Identities = 63/80 (78%), Positives = 73/80 (91%)

Query:     1 KVFVPEDQKDDKYWQRRKKNNVAAKRSRDARRLKENQITVRAAFLERENSALRQEVAELR 60
             KVFVP++QKD+KYW RRKKNNVAAKRSRDARRLKENQIT+RAAFLE+EN+ALR EVAELR
Sbjct:   182 KVFVPDEQKDEKYWTRRKKNNVAAKRSRDARRLKENQITIRAAFLEKENTALRTEVAELR 241

Query:    61 KDFGRCKNTVARYEAKYGAL 80
             K+ G+CK  V++YE KYG L
Sbjct:   242 KEVGKCKTIVSKYETKYGPL 261
```

Fig. 5 Deduced amino acid sequence and homology search. A. Deduced amino acid sequences from 3 possible reading frames shows that frame 3 encodes an open reading frame. B. Protein database search using frame 3 by BLAST program reveals high homology with PAR proteins, 80% identity and 93% similarity with chicken TEF homologue VBP. Query is the deduced peptide sequence of the 1.9 kb clone, while subject is the chicken VBP sequence.

We named the clone *tef* for the zebrafish homologue of the TEF protein. This clone was designated as *tef*1.9. Alignment of this polypeptide with all the identified PAR bZIP proteins reveals that it encodes DNA binding domain and leucine zipper domain characteristic of bZIP proteins (Fig. 1). In the leucine zipper domain it has three sequential leucine residues each of which is separated by six other residues. The heptad repeats have been shown to be critical for dimerization and dimerization specificity (Vinson et al, 1993). In the DNA binding domain it contains predominantly basic amino acid residues, in *tef* it contains 11 basic residues over the 25- residue polypeptide. Those two domains are considered conserved at high degree among the bZIP proteins from different species. Our observation is in agreement with this point.

3.2 Screening of an adult zebrafish cDNA library identified the full length *tefα*

At the beginning, we used the *EcoRI* fragment (1.2 kb) from the 1.9 kb cDNA clone, which included the putative DNA binding domain and leucine zipper domain and around 1 kb of 3' UTR, as probe labeled by α-^{32}P -dCTP to screen an embryonic zebrafish cDNA library. No positive clones were identified from the embryonic cDNA library. Considering the rat TEF was expressed exclusively at the pituitary region at day 14 during embryogenesis and widely expressed in adult tissues (Drolet et al, 1991), we reasoned that the negative screening was due to low expression level of this transcript in embryonic zebrafish. We switched to screen an adult zebrafish cDNA library using the same probe. In

the first round of screening we pulled out 26 positive plagues from 4.5×10^5 phages. The positive plagues were then subjected to second round of screening, 8 of them still showed positive signals. PCR, using the phagemids as template and T3 as forward primer and TEFA1 as reverse primer which is specific to 5' end of the 1.9 kb cDNA clone, revealed that 4 of the 8 positive plagues had 1.0 kb insert, 1 had 0.75 kb insert, and 3 had no PCR products (data not shown). It indicated that there were four clones containing inserts around 2.8 kb. Plasmid rescue and enzyme digestion confirmed the insert size around 2.8 kb(data not shown). One of the four clones was subjected for further analysis. It is designated as pBS*tef*2.8.

Sequencing subclones according to the restriction mapping and nest deletion subclones revealed the whole transcript (Fig. 6). The 2.7 kb cDNA transcript includes one open reading frame encoding a polypeptide of 300 amino acid residues. Translation start site is located at 174th nucleotide. The juxtaposed sequence, gacATGA, is relatively favorable compared to the Kozak consensus, pu(-3)pypyATGpu(+4) (here Pu means purine, while Py means pyrimidine), where -3 and +4 positions are important factors to determine the translation start site of a transcript (Kozak, 1984; Kozak, 1989; Grunert and Jackson, 1994). The most conserved and most important nucleotide at -3 is a purine (G). Furthermore, three in-frame stop codons (21-23, 65-67, and 101-103 in Fig. 6) are located upstream to this initiation site (Kozak, 1989; Grunert and Jackson, 1994). This cDNA contains a long stretch of 3' UTR. There is a non-canonical polyadenylation signal located at 16 bps upstream to the poly(A) tail (Fig. 6). Protein coding region (903 bps) shows

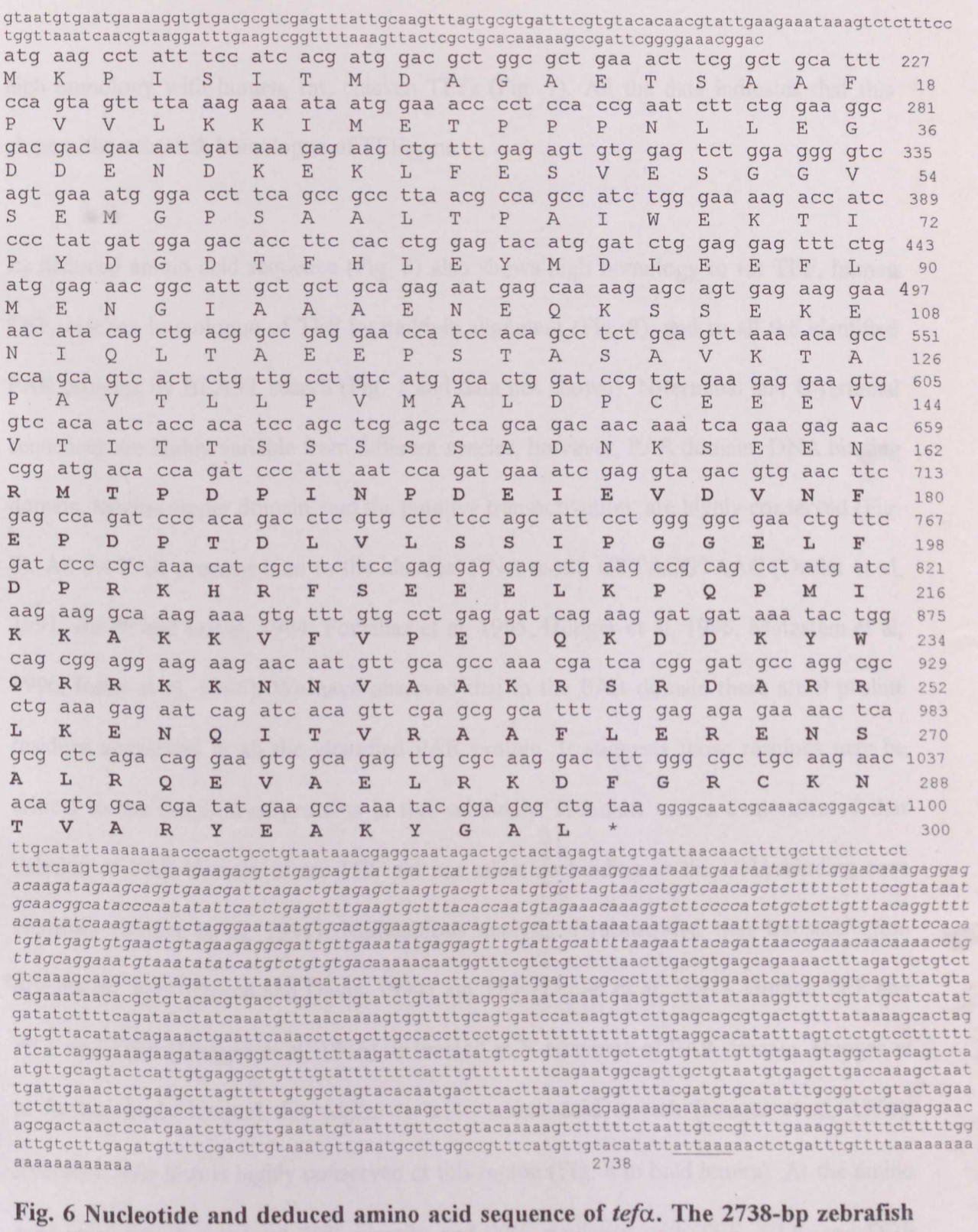

Fig. 6 Nucleotide and deduced amino acid sequence of *tef*α. The 2738-bp zebrafish *tef* cDNA sequence inclusive of the poly A track has been determined from overlapping cDNA clones. The region of characterized sequence of the PAR subfamily proteins, including PAR domain, DNA binding domain, and leucine zipper region, is hilighted in bold. The putative polyadenylation signal is underlined. The italic region is the 3' UTR region of *tef*β.

high homology with human, rat, chicken TEFs (Fig. 7). All the data indicates that this clone is the zebrafish homologue of TEF gene.

Its deduced amino acid sequence (Fig. 6) also shows high homology to rat TEF, human TEF, chicken homologue of TEF by multiple alignment (Fig. 8), and to all the identified PAR proteins by BLAST search (Fig. 1 and data not shown). N-terminal and C-terminal sequences are highly variable from different species, however, PAR domain, DNA binding domain, leucine zipper domain, and the putative transactivation are highly conserved (Fig. 8). All the PAR proteins bind to the identical DNA motif, GTTACGTAAC (Drolet et al, 1991; Burch and Davis, 1994; Fonjallaz et al, 1995; Hunger et al, 1996; Metzstein et al, 1996; Inaba et al, 1996). We have observed that in the PAR domain there are 9 proline residues conserved in all the identified PAR protein. It suggests those residues may be critical to the function of proteins in this subfamily. Gel shift assays demonstrated that deletion of PAR domain of rat TEF did not affect its binding to the DNA binding core elements (Drolet et al, 1991). Mutagenesis analysis demonstrated that transactivation domain of VBP (Burch and Davis, 1994), rat TEF (Drolet et al, 1991), Human TEF and HLF and DBP (Hunger et al, 1996), is located at N-terminus of PAR domain. Hunger et al further finely located this region to a 40 amino acid stretch (hunger et al, 1996). Zebrafish *tefα* also is highly conserved at this region (Fig. 8 in bold letters). At the amino acid level *tefα* has around 70% identity and 90% similarity with TEF, HLF, and DBP transcripts (data not shown). As Burch and Davis pointed out, even though the whole

```
 Sequences producing High-scoring Segment Pairs:                Score  P(N)       N
gb|S58745|S58745    thyrotroph embryonic factor=leucine z...  1086  2.2e-110   2
gb|U06935|HSU06935  Human thyrotroph embryonic factor (TE...  1068  8.7e-108   2
gb|U44059|HSU44059  Human thyrotroph embryonic factor (TE...  1068  2.7e-107   2
gb|U09221|GGU09221  Gallus gallus White Leghorn beta/beta...   995  1.5e-98    2

gb|S58745|S58745 thyrotroph embryonic factor [rats, pituitary, mRNA, 817 nt]
 Score = 1086 (300.1 bits), Expect = 2.2e-110, Sum P(2) = 2.2e-110
 Identities = 310/426 (72%), Positives = 310/426 (72%), Strand = Plus / Plus

Query:   478 GAAGAGAACCGGATGACACCAGATCCCATTAATCCAGATGAAATCGAGGTAGACGTGAAC 537
             ||| || |  ||  |||||||  ||||||  | ||  |     | |||||  || ||||||
Sbjct:   392 GAAAAGGAGAGGGAGACACCAAGTCCCATCGACCCCAACTGTGTGGAGGTTGATGTGAAC 451

Query:   538 TTCGAGCCAGATCCCACAGACCTCGTGCTGTCCAGCATTCCTGGGGGCGAACTGTTCGAT 597
             ||| | || || ||  | |||||  || || || ||  | || || || |||||  ||| |
Sbjct:   452 TTCAATCCTGACCCTGCTGACCTGGTCCTCTCTAGTGTACCAGGTGGGGAACTTTTCAAC 511

Query:   598 CCCCGCAAACATCGCTTCTCCGAGGAGGAGCTGAAGCCGCAGCCTATGATCAAGAAGGCA 657
             || || || ||    ||  | |||||  || ||||||||  ||||||||||||||||  ||
Sbjct:   512 CCTCGGAAGCACAAGTTTGCAGAGGAAGACCTGAAGCCCCAGCCTATGATCAAGAAAGCC 571

Query:   658 AAGAAAGTGTTTGTGCCTGAGGATCAGAAGGATGATAAATACTGGCAGCGGAGGAAGAAG 717
             ||||||||  |||||  |||||  ||  ||||||||||  ||  ||||||    ||  | ||||||
Sbjct:   572 AAGAAAGTCTTTGTCCCTGATGAGCAGAAGGATGAGAAGTACTGGACAAGGCGTAAGAAG 631

Query:   718 AACAATGTTGCAGCCAAACGATCACGGGATGCCAGGCGCCTGAAAGAGAATCAGATCACA 777
             ||||||||  |||||  |||||  ||  |||||  ||||||||||||||  ||||||||||||||
Sbjct:   632 AACAATGTGGCAGCTAAACGCTCCCGGGACGCCAGGCGCCTGAAGGAGAATCAGATCACC 691

Query:   778 GTTCGAGCGGCATTTCTGGAGAGAGAAAACTCAGCGCTCAGACAGGAAGTGGCAGAGTTG 837
              | ||  ||  ||||||||||||||  ||  |||  ||||  ||  |   |||  ||  ||  ||| |
Sbjct:   692 ATCCGGGCAGCATTTCTGGAGAAGGAGAACACAGCCCTGCGGACGGAGGTTGCTGAGCTT 751

Query:   838 CGCAAGGACTTTGGGCGCTGCAAGAACACAGTGGCACGATATGAAGCCAAATACGGAGCG 897
             ||||||||  | ||    |||||||  ||  |||  |    |||||  ||||  |||||  |
Sbjct:   752 CGCAAGGAGGTGGGCAAGTGCAAGACCATCGTGTCCAAGTATGAGACCAAGTACGGGCCC 811

Query:   898 CTGTAA 903
              |||||
Sbjct:   812 TTGTAA 817

 Score = 420 (116.1 bits), Expect = 2.2e-110, Sum P(2) = 2.2e-110
 Identities = 116/156 (74%), Positives = 116/156 (74%), Strand = Plus / Plus

Query:   135 GTTTGAGAGTGTGGAGTCTGGAGGGGTCAGTGAAATGGGACCTTCAGCCGCCTTAACGCC 194
             | | |||   |  |||||||  ||  | ||||   ||||     ||  |||  ||  | | |||
Sbjct:    85 GCTGGAGGAAGACGAGTCTGCAGCAGCCAGTACCATGGCCGTCTCTGCCTCCCTCATGCC 144

Query:   195 AGCCATCTGGGAAAAGACCATCCCCTATGATGGAGACACCTTCCACCTGGAGTACATGGA 254
             | ||||||||||  |||||  |||||||  |||||  ||  | ||||||||  ||||||||||
Sbjct:   145 ACCCATCTGGGACAAGACAATCCCCTACGATGGCGAGTCTTTCCACCTAGAGTACATGGA 204

Query:   255 TCTGGAGGAGTTTCTGATGGAGAACGGCATTGCTGC 290
              |||||  |||||  |||  |||||||  |||||  ||||
Sbjct:   205 CCTGGATGAGTTCCTGTTGGAGAATGGCATCCCTGC 240
```

Fig. 7 BLAST nucleotide search demonstrates high homology between *tef*α and human, rat, chicken (VBP) TEFs. Query is the *tef*α sequence, while subject is the rat TEF sequence.

```
1 HTEF_261_ ---------- ---------- ---------- -----MEN--PPREA RLDEEKGKEKLEEDE A-AAASTMA VSASLM     37
2 HTEF_303_ --------MSDAGGG KKPPVDPQAGPGPGP GRAAGERGLSGSFPL VLKKLMEN--PPREA RLDKEKGKEKLEEDE A-AAASTMA VSASLM     79
3 RTEF_261_ ---------- ---------- ---------- -----MEN--PPRET RLDKEKGKEKLEEDE S-AAASTMA VSASLM     37
4 VBPβ_293_ ---------- -----MSVCNSAGGP A--------ALDFPE VLKSLLEYSLPWTTK MTDKEK-KIKLEEDE A-AAASTMA VSASLM     60
5 VBPα_313_ MPGRAAHQEAAAAGG AAAEPTAAGGSAGAV AQQPEQQGLAGAFPL VLKKLMEN--PPRDA RLDKEK-KIKLEEDE A-AAASTMA VSASLM     86
6 tefβ_293_ ---------- ---------- --------MSSEIPE IFKALLEYP-FSLPS IDDNENDKEKLFESV ESGGVSEMG PSAALT     51
7 tefα_300_ ---------- ----------MKPIS ITMDAGAETSAAFPV VLKKIMETP-PPNLL EGDDENDKEKLFESV ESGGVSEMG PSAALT     64

1 HTEF_261_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPASP THLAHN--LL LPVAELE-----GKE SASSSTASPPSSS-- ---TAIFQPSETVSS    115
2 HTEF_303_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPASP THLAHN--LL LPVAELE-----GKE SASSSTASPPSSS-- ---TAIFQPSETVSS    157
3 RTEF_261_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPASP THLAQN--LL LPVAELE-----GKE SASSSTASPPSSS-- ---TAIFQPSETVSS    115
4 VBPβ_293_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPSSP THLDLNQNPL MPVAKLE-----EKE PASASTGSPVSSSS- ---TAVYQQSEAASS    141
5 VBPα_313_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPSSP THLDLNQNPL MPVAKLE-----EKE PASASTGSPVSSSS- ---TAVYQQSEAASS    167
6 tefβ_293_ PAIWEKTIPYDGDTF HLEYMDLEEFLMENG IAAAE NEQKSSEKEN IQLTAEEPSTASAVK TAPAVTLLPVMALDP CEEEVVTITTSSSSS    141
7 tefα_300_ PAIWEKTIPYDGDTF HLEYMDLEEFLMENG IAAAE NEQKSSEKEN IQLTAEEPSTASAVK TAPAVTLLPVMALDP CEEEVVTITTSSSSS    154

1 HTEF_261_ TESSLEKERETPSPI DPNCVEVDVNFNPDP ADLVLSSVPGGELFN PRKHKFAEEDLKPQP MIKKAKKVFVPDEQK DEKYWTRRKKNNVAA    205
2 HTEF_303_ TESSLEKERETPSPI DPNCVEVDVNFNPDP ADLVLSSVPGGELFN PRKHKFAEEDLKPQP MIKKAKKVFVPDEQK DEKYWTRRKKNNVAA    247
3 RTEF_261_ TESSLEKERETPSPI DPNCVEVDVNFNPDP ADLVLSSVPGGELFN PRKHKFAEEDLKPQP MIKKAKKVFVPDEQK DEKYWTRRKKNNVAA    205
4 VBPβ_293_ TESPPQNERNTPSPI DPDCVEVEVNFNPDP ADLVLSSVPGGELFN PRKHKFTEEDLKPQP MIKKAKKVFVPDEQK DEKYWTRRKKNNVAA    231
5 VBPα_313_ TESPPQNERNTPSPI DPDCVEVEVNFNPDP ADLVLSSVPGGELFN PRKHKFTEEDLKPQP MIKKAKKVFVPDEQK DEKYWTRRKKNNVAA    257
6 tefβ_293_ ADNKSEENRMTPDPI NPDEIEVDVNFEPDP TDLVLSSIPGGELFD PRKHRFSEEELKPQP MIKKAKKVFVPEDQK DDKYWQRRKKNNVAA    231
7 tefα_300_ ADNKSEENRMTPDPI NPDEIEVDVNFEPDP TDLVLSSIPGGELFD PRKHRFSEEELKPQP MIKKAKKVFVPEDQK DDKYWQRRKKNNVAA    244

1 HTEF_261_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGKCKTI VSKYETKYGPL---- --    261
2 HTEF_303_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGKCKTI VSKYETKYGPL---- --    303
3 RTEF_261_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGKCKTI VSKYETKYGPL---- --    261
4 VBPβ_293_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGRCKNI VSKYETRYGPFDLSD SE    293
5 VBPα_313_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGRCKNI VSKYETRYGPL---- --    313
6 tefβ_293_ KRSRDARRLKENQIT VRAAFLERENSALRQ EVAELRKDFGRCKNT VARYEAKYGALGPEE DV    293
7 tefα_300_ KRSRDARRLKENQIT VRAAFLERENSALRQ EVAELRKDFGRCKNT VARYEAKYGAL---- --    300
```

Fig. 8 Multiple alignments of rat, human TEFs, and chicken VBPs. It shows highly variable N-terminal and C-terminal regions. The putative transactivation domain is in bold letters.

PAR domain of VBP did not have transacting ability in vitro, it may harbor inhibitory elements (Burch and Davis, 1994). Fine mapping of PAR domain and proline residues in PAR domain will reveal the real function of this domain.

In summary, from analysis of nucleotide and its deduced amino acid sequences, it is evident that the 2.8 kb zebrafish cDNA encodes homologue of TEF gene. It has the characteristics of PAR proteins, including C-terminal leucine zipper domain, basic DNA binding domain, proline and acidic rich domain, and transacivation domain. All those domains are conserved in all identified PAR proteins.

3.3 RACE identified zebrafish *tefβ*

While we were screening for the full-length *tefα* cDNA, a 0.8 kb fragment was identified by 5'RACE PCR using the adult cDNA library as template and reverse primer TEFA1 which is specific to the 5' end of the 1.9 kb cDNA as gene-specific primer (Wong, 1996). Sequencing data showed that this fragment encodes a protein with an identical C-terminus to *tefα* and a distinct N-terminus. Burch and Davis (1994) reported cloning and characterization of VBP isoforms with distinct N-terminus and/or C-terminus, and different tissue-specific expression. Those facts indicate there may be some other *tef* isoforms. Using SK and TEFA1 as primers and the adult cDNA library as template for first round PCR and T3 and TEFA1 as nested primers for nest PCR (PCR program, initiate denaturation at 94°C for 5 minutes, 30 cycles each of which is 94°C for 30'', 55°C

for 1', and 72°C for 1'), we successfully cloned two PCR fragments of 0.9 kb and 0.75 kb. Sequencing of 0.75 kb fragment demonstrated identical sequence to *tefα* (data not shown), and sequencing of 0.9 kb fragment revealed a distinct 5' end sequence from *tefα*, which is the same as the 0.8 kb clone identified by Wong (Wong, 1996), but contains additional 5' upstream sequences. We named the clone *tefβ*. Thus, at least two different *tef* isoforms exist in zebrafish.

3' RACE PCR using the adult cDNA library as template, TSK1 forward primer (specific to the 5' end of the 0.9 kb fragment), and T7 reverse primer (PCR program, initial denaturation for 5 minutes at 94°C followed by 30 cycles each of which included 94°C for 30'', 55°C for 1', and 72°C for 2'30'') identified a major product whose size was around 1.4 kb (data not shown). Sequencing of this 1.4 kb clone by gene-specific primers revealed the whole transcript of *tefβ*, which has a distinct C-terminus and a distinct N-terminus (Fig. 9 and underlined sequences there). Translation initiation site is located to 74th nucleotide whose juxtaposed sequence, gagATGT, has a relatively favorable environment, matching to Kozak consensus (Kozak, 1984; Kozak, 1989; Grunert and Jackson, 1994). It has purine (G) in -3 position but T instead of G in +4 position. Further more an in-frame stop codon (57-59 th nucleotide in Fig. 9) is located 12 bps upstream to this initiation site. It contains a poly(A) tail. However, there is no consensus polyadenylation signal. Instead, a non-canonical polyadenylation site, AATATA, is located at 15 nucleotides upstream to the poly (A) tail (1421-1427 in Fig. 9). It has a truncated 3' UTR compared to *tefα*. It encodes a distinct C-terminus and N-terminus, which do not have significant homology to

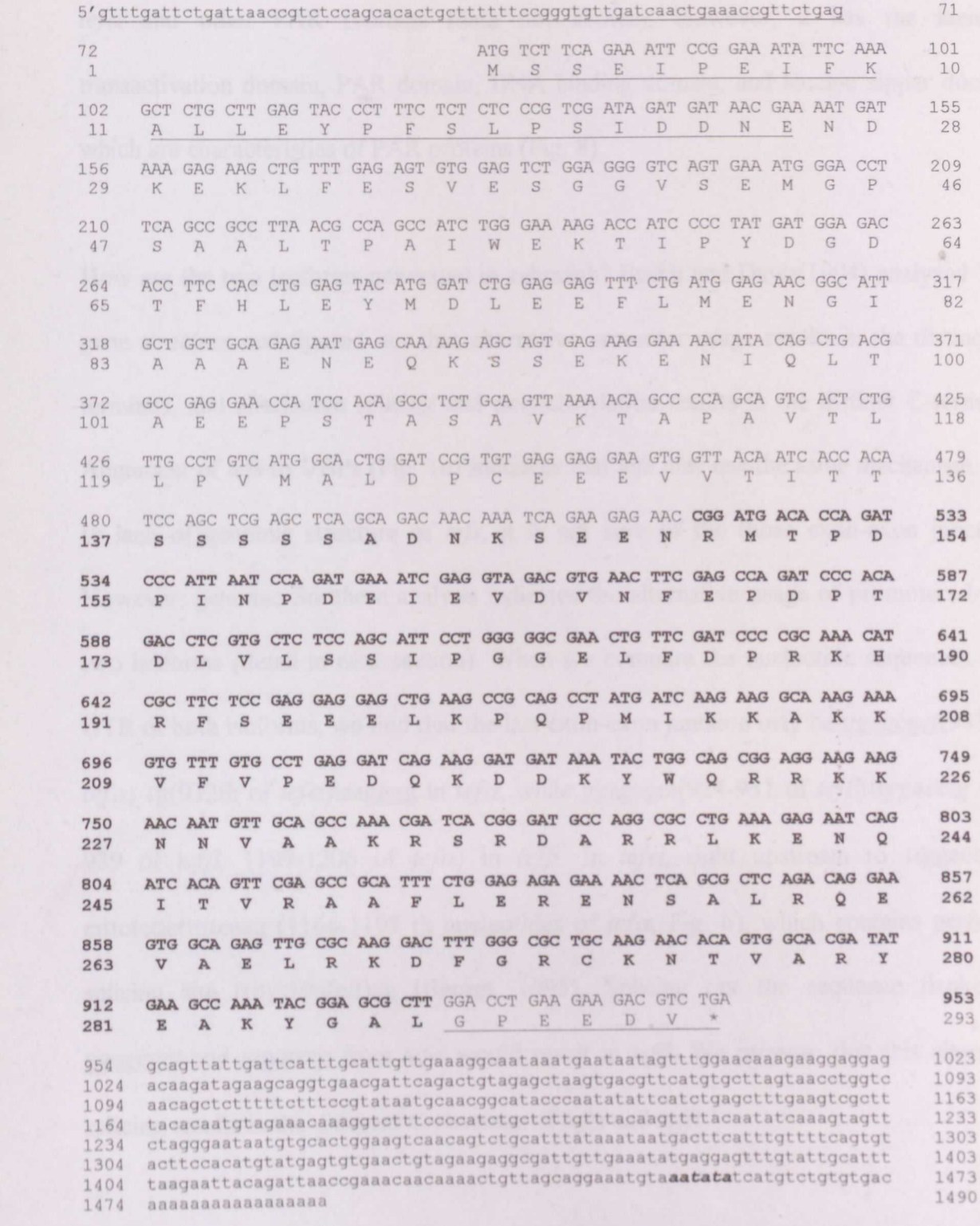

Fig. 9 Nucleotide and deduced amino acid sequence of *tef*β. The 1490-bp zebrafish *tef*β isoform sequence is determined from RACE PCR clones. Note that 5'end and 3'end of coding region which are different from these of *tef*α are underlined. The region of characterized sequence of PAR proteins, including PAR domain, DNA binding domain, and leucine zipper region , is in bold letters. The putative polyadenylation signal is in italic and bold letters.

tefα and other PAR proteins (data not shown). However, it has the identical transactivation domain, PAR domain, DNA binding domain, and leucine zipper domain, which are characteristics of PAR proteins (Fig. 8).

How are the two isoforms generated in zebrafish? Burch and Davis(1994) analyzed VBP gene structure and figured out that alternative promoter usage results in the distinct N-terminus, and alternative splicing and polyadenylation results in the distinct C-terminus. Alignment of *tefs* to VBPs (Fig. 10) indicates that *tefs* may use the same mechanism. Due to lack of genomic structure of *tefs*, it is not sure of the those exon-exon junctions. However, genomic Southern analysis indicates the alternative usage of promoters by the two isoforms (detail in next section). When we compare the nucleotide sequences of 3' UTR of both isoforms, we find that the last exon-exon junction may be *ggagcgct*(931st of *tefα*) ↓g(932th of *tefα*)taagggg in *tefα*, while ggagcgct(924-931 of *tefβ*)↓***tggacctg*** (932-939 of *tefβ*, 1199-1206 of *tefα*) in *tefβ*. In *tefα*, right upstream to ↓tggacctg is ctttctcttcttttttcaag (1164-1198 th nucleotides of *tefα*, Fig. 6), which contains perfect 3' splicing site ((py)10n(c/t)ag (Berget, 1995). Splicing out the sequence flanked by *ggagcgct* and *tggacctg* from *tefα* would result in *tefβ*. We propose that this alternative splicing results in the different C-terminus of *tefα* and *tefβ*.

3.4 Genomic Southern analysis indicates that *tef* is a single copy gene and the two isoforms may use alternative promoters

```
1 VBPβ_293_ --------------- -----MSVCNSAGGP A--------ALDFPE VLKSLLEYSLPWTTK MT↓DKE--KKIKLEED EAAAASTMAVSASLM     60
2 VBPα_313_ MPGRAAHQEAAAAGG AAAEPTAAGGSAGAV AQQPEQQGLAGAFPL VLKKLMEN--PPRDA RL↓DKE--KKIKLEED EAAAASTMAVSASLM     86
3 tefβ_293_ --------------- --------------- M--------SSEIPE IFKALLEYP-FSLPS ID↓DNENDKEKLFESV ESGGVSEMGPSAALT     51
4 tefα_300_ --------------- --MKPISITMDAGAE T--------SAAFPV VLKKIMETP-PPNLL EG↓DDENDKEKLFESV ESGGVSEMGPSAALT     64

1 VBPβ_293_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPSSPTHLDLNQNPL MPVAKLEE-----KE PASASTGSPVSSSS- ---TAVYQQSEAASS    141
2 VBPα_313_ PPIWDKTIPYDGESF HLEYMDLDEFLLENG IPSSPTHLDLNQNPL MPVAKLEE-----KE PASASTGSPVSSSS- ---TAVYQQSEAASS    167
3 tefβ_293_ PAIWEKTIPYDGDTF HLEYMDLEEFLMENG IAAAENEQKSSEKEN IQLTAEEPSTASAVK TAPAVTLLPVMALDP CEEEVVTITTSSSSS    141
4 tefα_300_ PAIWEKTIPYDGDTF HLEYMDLEEFLMENG IAAAENEQKSSEKEN IQLTAEEPSTASAVK TAPAVTLLPVMALDP CEEEVVTITTSSSSS    154

1 VBPβ_293_ TE↓SPPQNERNTPSPI DPDCVEVEVNFNPDP ADLVLSSVPGGELFN PRKHKFTEEDLKPQP MIKKAKKVFVP↓DEQK DEKYWTRRKKNNVAA    231
2 VBPα_313_ TE↓SPPQNERNTPSPI DPDCVEVEVNFNPDP ADLVLSSVPGGELFN PRKHKFTEEDLKPQP MIKKAKKVFVP↓DEQK DEKYWTRRKKNNVAA    257
3 tefβ_293_ AD↓NKSEENRMTPDPI NPDEIEVDVNFEPDP TDLVLSSIPGGELFD PRKHRFSEEELKPQP MIKKAKKVFVP↓EDQK DDKYWQRRKKNNVAA    231
4 tefα_300_ AD↓NKSEENRMTPDPI NPDEIEVDVNFEPDP TDLVLSSIPGGELFD PRKHRFSEEELKPQP MIKKAKKVFVP↓EDQK DDKYWQRRKKNNVAA    244

1 VBPβ_293_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGRCKNI VSKYETRYGP↓FDLSD SE    293
2 VBPα_313_ KRSRDARRLKENQIT IRAAFLEKENTALRT EVAELRKEVGRCKNI VSKYETRYGP↓L---- --    313
3 tefβ_293_ KRSRDARRLKENQIT VRAAFLERENSALRQ EVAELRKDFGRCKNT VARYEAKYGA↓LGPEE DV    293
4 tefα_300_ KRSRDARRLKENQIT VRAAFLERENSALRQ EVAELRKDFGRCKNT VARYEAKYGA↓L---- --    300
```

Fig. 10 Alignment of *tefs* to VBPs indicates that *tefs* may have the similar exon-exon junctions. ↓ shows exon-exon junction in VBPs and putative exon-exon junction in *tefs*.

Fifty microgram of adult zebrafish genomic DNA digested by appropriate restriction enzymes were separated by 1.0% agarose gel and transferred to a nylon membrane by standard capillary blotting. A ^{32}P labeled 2 kb probe of *tefα* including 5' UTR and ORF and partial 3' UTR (*EcoR I* fragment from pBS *tefα*) was hybridized to this blot (Fig. 11A); the identical blot was stripped and reprobed by a ^{32}P labeled 86 bp probe specific to *tefβ* covered 5' UTR (*BspE I* and *EcoR I* fragment from pBS *tefβ*) (Fig. 11B). Since 3'UTR of VBP transcript and human TEF transcript are within the last exon of the genes(Burch and Davis, 1994; Khatib et al, 1994), and in most cases the whole 3'UTR is within the last exon of a gene (Berget, 1995), we assume that the last exon of zebrafish *tef* would cover the 3'UTR. Therefore the *tef* 2.0 kb probe should cover the whole gene if it is used as a probe for genomic Southern analysis. Restriction mapping indicated that there is no *EcoR I* site, one *BamH I* site, and one *Pst I* site within this 2.0 kb fragment (data not shown). If zebrafish *tef* gene is a multiple-copy gene, this probe should detect more than one fragments in the *EcoR I* digested genome, more than two in the *BamH I* digested or *Pst I* digested genome. However, one fragment was detected in the *EcoR I* digested genome (lane 2 in Fig. 11A), and two fragments only were detected in the either *BamH I* digested or *Pst I* digested genome (Lane 1 and 5 in Fig. 11A). This signal patterns indicate that zebrafish *tef* gene is a single copy gene in the genome. Comparing the two signal patterns, we noticed that *tefβ* probe specific to the putative first exon detected several distinct bands (Fig.11B) from *tefα* probe (Fig.11A). For example, 8 kb band of the *EcoRI* digested genome, 15 kb band of the *EcoRV* digested genome, 2 kb and 3 kb bands of the *HindIII* digested genome (Fig. 11B). If *tefβ* were using the same promoter in the gene as

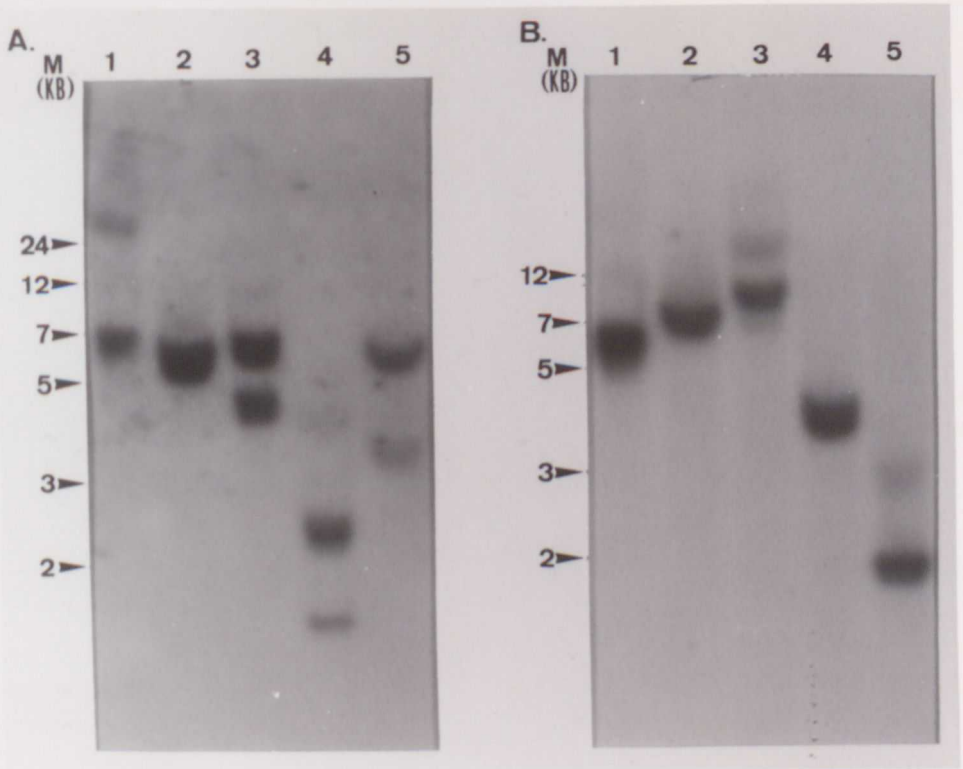

Fig. 11 Genomic Southern blot using 50 μg of zebrafish adult genomic DNA digested by *BamHI* (lane 1), *EcoRI* (lane 2), *EcoRV*(lane 3), *HindIII* (lane 4), and *PstI* (lane 5) was individually probed by (A). 2 kb fragment including 5' UTR, coding region, and partial 3'UTR of *tefα* or (B). 86 bp fragment of 5'UTR of *tefβ*. Numbers in the left of picture show DNA molecular weight in kb. The films were exposed at -70°C for three days.

tefα , the signal patterns by *tefα* probe including *tefα* 5' UTR would have included every signal detected only by *tefβ* 5' UTR probe. The fact that the *tefβ* probe gives different bands compared to *tefα* probe indicates that the two isoforms may use different promoters. Usage of alternative promoters was observed in chicken VBPs (Burch and Davis, 1994) and rat HLFs (Fonjallaz et al, 1996). In both cases, the different isoform proteins show tissue-specific expression and target preference. Due to the different combination of C-terminus and N-terminus of VBP, VBP gene encodes at least four transcripts. It is reasonable to postulate that there may be four different transcripts encoded by zebrafish *tef* gene. Since rat TEF is expressed exclusively at the developing anterior pituitary gland during embryogenesis and ubiquitously in adult tissues, it is interesting to contemplate whether one isoform of *tefs* is specific to pituitary and targets TSHβ, while the other one is specific for adult tissues.

3.5 Northern analysis revealed the single transcript of 2.8 kb corresponding to zebrafish *tefα*

Northern blot using 30 μg of zebrafish adult total RNA by 5'UTR probe (1 to 236 bp of *tefα*, data not shown), *tefα* probe (2.0 kb fragment from 1 to 2074 bp, data not shown), and 3' UTR (0.8 kb fragment from 1522 to 2358 bp of *tefα*, result in Fig.11A) revealed the single transcript of around 2.8 kb, indicating that the *tefα* cDNA clone is almost full-length. However, the message that corresponds to *tefβ* can not be detected in this assay, probably it is due to the very low transcription level.

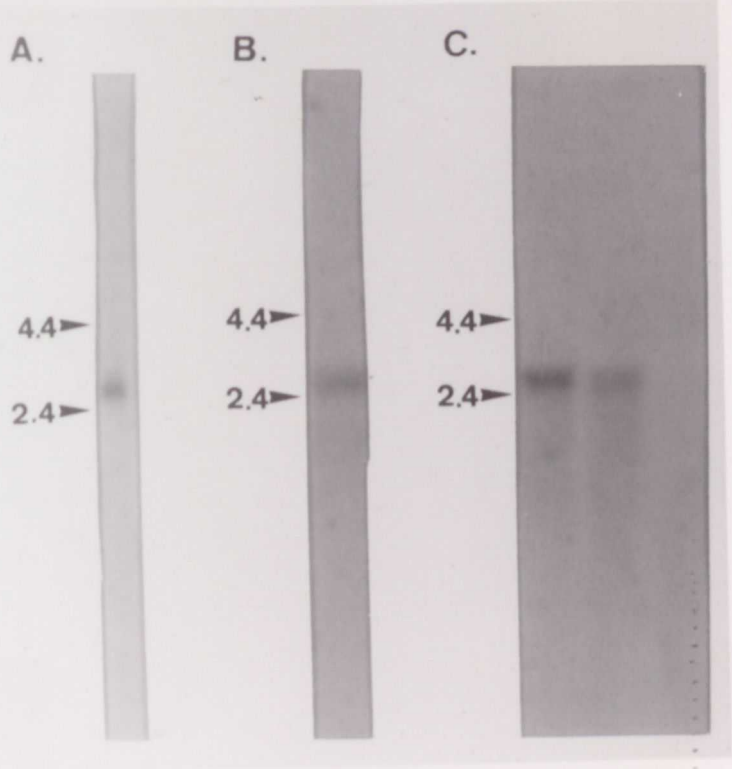

Fig. 12 Northern blot using 30 μg of zebrafish total RNA were hybridized with *tefα* 3'UTR probe (*HindIII-HincII* fragment , 1522 to 2358 bp of *tefα*). Exposure was under -70°C with intensifiers for three days. Numbers in the left of picture indicate RNA molecular weight in kb. A single 2.8 kb transcript was detected. A. adult RNA. B. 4-day RNA. C. from left to right, high-pec stage RNA, Prim-22 stage RNA, and Prim-6 stage RNA.

Rat TEF transcript starts to express on day 14 during embryogenesis and is exclusively expressed at the developing anterior pituitary gland. We are interested to know from which stage of embryogenesis *tef* transcripts are expressed. Thirty μg of different stages (high-pec, prim-22, and prim-6 stages) of embryo total RNA probed by the 3' UTR fragment showed that expression of *tefα* could be first detected at prim-22 stage (35 hpf) during development (Fig. 12B and C). However, due to the usage of total RNA instead of poly(A)+ RNA, we can not exclude the possibility that *tefα* transcript is expressed earlier. In addition whole-mount embryo in situ hybridization would be helpful to localize the transcripts during embryogenesis.

3.6 RT-PCR detected both transcripts

Because *tefβ* message was not detected in Northern blot analysis, we decided to switch to more sensitive method, RT-PCR. A zebrafish β-actin cDNA clone (from Dr. Z. GONG, School of Biological Sciences, National University of Singapore) was used as an internal control. Two primers were synthesized from the zebrafish β-actin, which flank 560 bps of β-actin transcript. Specific primers for *tefα* (TEFAS and TEFA1) flanks 770 bps of *tefα*, and specific primers for *tefβ* (TEFSK2 and TEFA1) will yield a 790 bp PCR product. RT-PCR program we used was reverse transcription at 48°C for one hour, followed by PCR (inactivation of reverse transcriptase and initial denaturation at 94°C for 2 minutes, 15-40 cycles each of which included denaturation at 94°C for 30", annealing at 60°C for 1 minute, and extension at 68°C for 2 minutes, and followed by one cycle of final extension

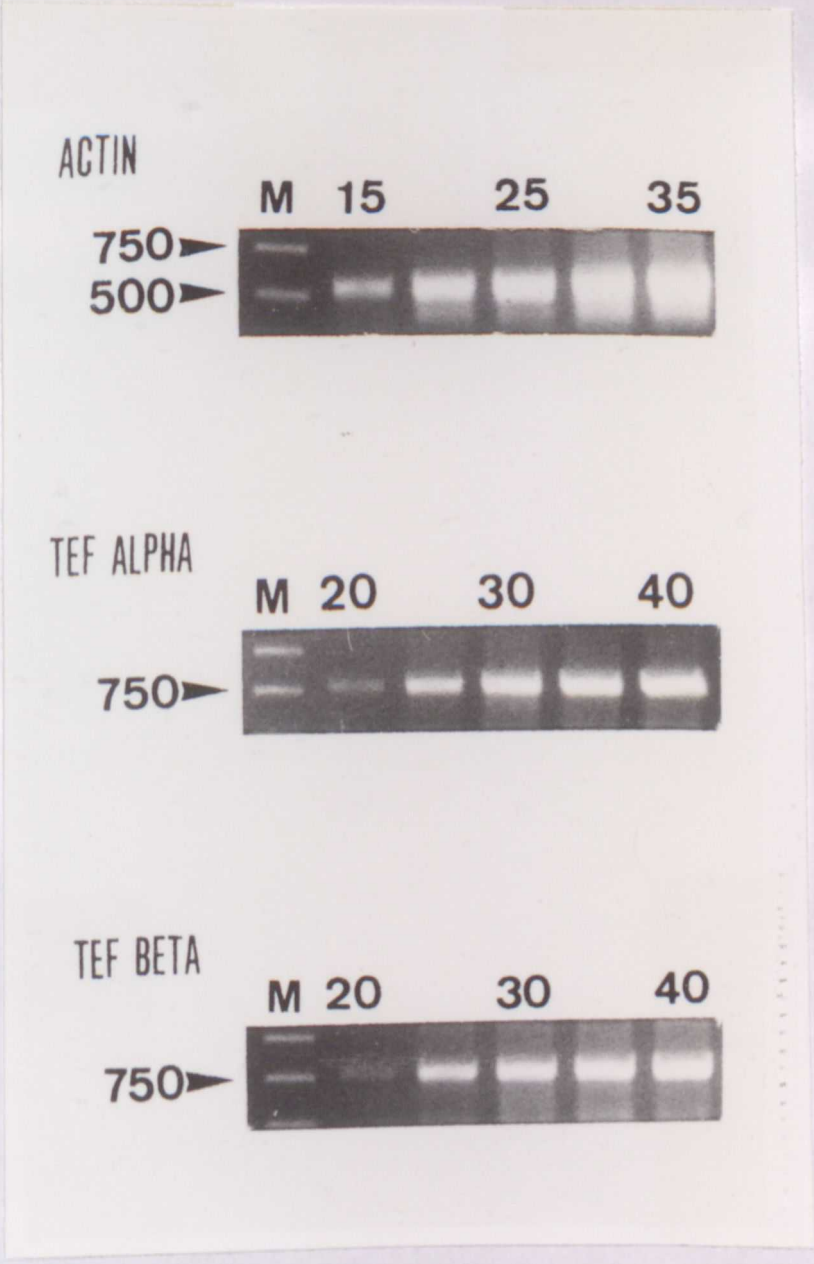

Fig. 13 RT-PCR demonstrated the expression of the two isoforms. The identical master mix except the different set of primers was used to amplify β actin transcript (top panel), *tef*α transcript (middle panel), *tef*β transcript (bottom panel) from mouth-stage total RNA by RT-PCR access system (Promega). Titration of PCR cycles was performed, 10 μl aliquot from each time point was collected and kept at -20°C, all samples were electrophoresized on 1.2% agarose gel. Numbers at the left side of pictures show the DNA marker in bp, at the top show PCR cycles.

at 68°C for 7 minutes). Fig. 13 shows that all three sets of primers can amplify the expected PCR fragments. It is a little bit surprising that RT-PCR results showed that there was not much difference between *tefα* and *tefβ* of RT-PCR products though *tefβ* transcript could not be detected in Northern analysis while *tefα* transcript could. However, since PCR exponentially amplifies DNA fragments, we will need quantitative RT-PCR for relative quantitation. At least RT-PCR demonstrated that there are two species of *tef* transcripts.

3.7 Possible functions of the zebrafish *tef*s

Homology comparison at the nucleotide level and amino acid level of the two cDNA clones has revealed that they encode two transcripts of zebrafish *tef*. Based on the functional analysis of human and rat TEF and chicken VBP, we propose function of zebrafish *tef*s. It is well known that all PAR proteins have the similar DNA binding domain, and transactivation domain. They preferentially self dimerize or with a member of PAR proteins. Fine mapping using a series of deletion mutant to drive the appropriate reporter constructs would reveal characteristics of zebrafish *tef*s (domain mapping, heterodimerization preference, and target preference). Genomic structure of this gene would confirm alternative usage of promoters and alternative splicing for the two isoforms. Since function of PAR domain is unknown, it has been proposed that it harbors both activation and inhibition domain (Drolet et al, 1991), fine mapping of this region

would be intriguing. Since the nine proline residues in PAR domain are highly conserved, point mutation of those residues would further elucidate function of this domain.

It is well established that rat TEF transcript starts to show up exclusively at the anterior pituitary on day 14 embryo, and transactivate thyroid stimulating hormone β subunit, and is widely expressed in most adult tissues (Drolet et al, 1991). Human TEF is also widely expressed in most tissues with relatively high level in kidney and liver and lymphoid cell line (Khatib et al, 1994). Chicken VBP is also ubitiquously expressed. However, expression of isoforms differs in different tissues. VBP can also bind to vitellogenin II promoter and transactivate it (Burch and Davis, 1994; Iyer et al, 1991). It suggests that zebrafish *tef* would be widely expressed in most of adult tissues. During embryogenesis, one isoform may turn on genes in thyrotroph cells.

Recently Metstein et al (1996) identified a negative transcriptional regulator of programmed cell death encoded by C. elegans gene CES2. Its DNA binding domain and leucine zipper domain are most similar to members of PAR proteins. CES2 protein has the same DNA binding specificity as PAR proteins. Inaba et al (1996) demonstrated that dominant-negative E2A-HLF chimaeric transcription factor can function to reverse apoptosis. Those facts suggest PAR proteins may be involved in apoptosis. Fig. 3 shows the comparison of CES2 and *tefβ* by BLAST program, both have conserved DNA binding domain and leucine zipper domain. Furthermore, partial PAR domain of *tefβ* (36 residues over 50 residues) has conservativity with CES2 right upstream of its DNA binding

domain. Exploration of DNA binding domain of *tef* and cooperation of PAR domain and DNA binding domain of *tef* in regulation of apoptosis would be very interesting.

3.8 Conclusion

Two cDNA clones have been isolated from zebrafish (Danio rerio). Sequence homology search revealed that they encode two isoforms of zebrafish *tef* transcripts with distinct N-terminus and C-terminus, designated as *tefα* (GenBank accession number:U43671)and *tefβ* (GenBank accession number:U96848) Both have highly conserved DNA binding domain, leucine zipper domain, and PAR domain compared to all the identified PAR proteins. Sequence analysis suggests that alternative splicing results in the different C-terminus, while Southern analysis indicates that the different N-terminus results from alternative usage of promoters. Northern analysis reveals single transcript around 2.8 kb, corresponding to the full-length *tefα*, which can be first detected at the stage of prim-22 during embryogensis by this assay. RT-PCR demonstrates the existence of both transcripts. Fine mapping of the transactivation domain and mutagenesis research on the PAR domain would be interesting future experiments. In addition, *tef* shows homology with CES2 gene in the DNA binding domain, leucine zipper domain, and PAR domain. Since CES2 gene is a negative regulator of programmed cell death, there is a possibility that *tef* is also involved in apoptosis.

4. REFERENCES

Arias, J., Alberts, A.S., Brindle, P., Claret, F.X., Smeal, T., Karin, M., Feramisco, J., and Montminy, M. (1994). Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. **Nature 370**, 226-229.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. **Journal of Biological Chemistry 270**, 2411-2414.

Burch J.B.E., and Davis, D. (1994). Alternative promoter usage and splicing options result in the differential expression of mRNA encoding four isoforms of chicken VBP, a member of the PAR subfamily of bZIP transcription factors. **Nucleic Acid Research 22**, 4733-4741.

Calkhoven, C.F., and Geert, A.B. (1996). Multiple steps in the regulation of transcription-factor level and activity. **Biochemical Journal 317**, 329-342.

Cereghini, S. (1996). Liver-enriched transcription factors and hepatocyte differentiation. **FASEB Journal** 10, 267-282.

Chomczynski, P., and Sacchi, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Analytical Biochemistry 162**, 156-159.

Chomczynski, P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. **BioTechniques 15**, 532-537.

Drolet, D.W., Scully, K.M., Simmons, D.M., Wegner, M., Chu, K., Swanson, L.W., and Rosenfeld, G. (1991). TEF, a transcription factor expressed specifically in the anterior

pituitary during embryogenesis, defines a new class of leucine zipper proteins. **Genes & Development 5**, 1739-1753.

Eisen, S.J. (1997). Zebrafish make a big splash. **Cell 87**, 969-967

Ellenberger, T. (1994). Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. **Current Opinion in Structural Biology 4**, 12-21.

Falvey, E., Fleury-Olela, F., and Schibler, U. (1995). The rat hepatic leukemia factor (HLF) gene encodes two transcriptional activators with distinct circadian rhythms, tissue distributions and target preferences. **EMBO Journal 14**, 4307-4317.

Fonjallaz, P., Ossipow, V., Wanner, G., and Schibler, U. (1996). The two PAR leucine zipper proteins, TEF and DBP, display similar circadian and tissue-specific expression, but have different target promoter preferences. **EMBO Journal 15**, 351-362.

Foulkes, N.S., and Saaaone-Corsi, P. (1992). More is better: activators and repressors from the same gene. **Cell 68**, 411-414.

Grunert, S., and Jackson, R.J. (1994). The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. **EMBO Journal 13**, 3618-3630.

He, X., and Rosenfeld, G. (1991). Mechanisms of complex transcriptional regulation: implications for brain development. **Neuron 7**, 183-196.

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? **Genes & Development 7**, 1291-1308.

Hunger, S.P., Ohyashiki, K., Toyama, K., and Cleary, M.L. (1992). Hlf, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. **Genes & Development 6**, 1608-1620.

Hunger, S.P., Li, S., Fall, M.J., Naumovski, L., and Cleary, M. (1996). The pro-oncogene HLF and the related basic leucine zipper protein TEF display highly similar DNA-binding and transcriptional properties. **Blood 87**, 4607-4617.

Inaba, T., Roberts, W.M., Shapiro, L.H., jolly, K.W., Raimondi, S.C., Smith, S.D., and Look, A.T. (1992). Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. **Science 257**, 531-534.

Inaba, T., Shapiro, L.H., Funabiki, T., Sinclair, A.E., jones, B.G., Ashmun, R.A., and Look, A.T. (1994). DNA-binding specificity and trans-activating potential of the leukemia-associated E2A-hepatic leukemia factor fusion protein. **Molecular and Cellular Biology 14**, 3403-3413.

Inaba, T., Inukai, T., Yoshihara, T., Seyschab, H., Ashmun, R.A., Canman, C.E., Laken, S.J., Kastan, M.B., and Look, A.T. (1996). Reversal of apoptosis by the leukaemia-associated E2A-HLF chimaeric transcription factor. **Nature 382**, 541-544.

Iyer, S.V., Davis, D.L., Seal, S.N., and Burch, J.B.E. (1991). Chicken vitellogenin gene-binding protein, a leucine zipper transcription factor binds to an important control element in the chicken vitellogenin II promoter, is related to rat DBP. **Molecular and Cellular Biology 11**, 4863-4875.

Kamei, Y., Xu,, L., Heizel, T., Torchia, J., Kurokawa, EGFR., Gloss, B., Lin, S., Heyman, R.A., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. **Cell 85**, 403-414.

Khatib, Z.A., Inaba, T., Valentine, M., and Look, A.T. (1994). Chromosomal localization and cDNA cloning of the human DBP and TEF genes. **Genomics 23**, 344-351.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. **Nucleic Acid Research 12**, 857-872.

Kozak, M. (1989). The scanning model for translation: an update. **Journal of Cell Biology** 108, 229-241.

Kwok, R.P.S., Lundblad, J.R., Richards, J.P., Bachinger, H.P., Brennan, R.G., Roberts, S.G.E., Green, M.R., and Goodman, R.H. (1994). Nuclear protein CBP is a coactivator for the transcription factor CREB. **Nature 370**, 223-226.

Landschultz, W., Johnson, W., and McKnight, S. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. **Science 243**, 1759-1764.

Lavery, D.J., and Schibler, U. (1993). Circadian transcription of the cholesterol 7α hydroxylase gene may involve the liver-enriched bZIP protein DBP. **Genes & Development 7**, 1871-1884.

Lewin, B. (1994). **Genes V** 5th edition. Oxford University Press, Oxford, London.

Marks, D.D., Marks, A.D., and Smith, C.M. (1996). **Basic Medical Biochemistry**. Williams & Wilksons Inc., Baltimore, Maryland.

Metzstein, M.M., Hengartner, M.O., Tsung, N., Ellis, R.E., and horvitz, H.R. (1996). Transcriptional regulator of programmed cell death encoded by Caenorhabditis elegans gene CES-2. **Nature 382**, 545-547.

Muelleer, C., Maire, P., and Schibler, U. (1990). DBP, a liver-enriched transcriptional activator is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. **Cell 61**, 279-291.

Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. **Nature 287**, 795-801.

O'Shea, E.K., Rutkowski, EGFR., and Kimp, S. (1992). Mechanism of specificity in the fos-jun oncoprotein heterodimer. **Cell 68**, 699-708.

Pabo, C.O. (1992). Transcription factors: structural families and principles of DNA recognition. **Annual Review Biochemistry 61**, 1053-1095.

Poslethwait, J.H., Jonson, S.L., Midson, C.N., Talbot, W.S., Gates, M., Ballinger, E.W., Africa, D., Andrews, EGFR. Carl, T., Eisen, J.S., Horne, S., Kimmel, C.B., Hutchinson, M., Jonson, M., and Rodriguez, A. (1994). A genetic linkage map for the zebrafish. **Science 264**, 699-703.

Ranish, J.A., and Hahn, S. (1996). Transcription: basal factors and activation. **Current Opinion in Genetics & Development 6**, 151-158.

Rossant, J., and Hopkins, N. (1992). Of fin and fur: mutational analysis of vertebrate embryonic development. **Genes & Development 6**, 1-13.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). **Molecular Cloning** 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. **Proceedings of National Academy of Sciences of USA 74**, 5463-5467.

Schuermann, M., Hunter, J.B., Hennig, G., and Muller, R. (1991). Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization. **Nucleic Acids Research 19**, 739-746.

Streisinger, G., Walker, C., Dower, N., Knauber, D., and Singer, F. (1981). Production of clones of homozygous diploid zebrafish (*Branchydanio rerio*). **Nature 291**, 293-296.

Struhl, K. (1989). Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. **Trends in Biochemistry 14**, 137-140.

Tansey, W.P., and Herr, W. (1997). TAFs: guilt by association? **Cell 88**, 729-732.

van Holde, K., and Zlatanova, J. (1996). Chromatin architectural proteins and transcription factors: a structural connection. **BioEssay 18**, 697-700.

Vinson, C.R., Hai, T., and Boyd, S.M. (1993). Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. **Genes & Development 7**, 1047-1058.

Weinberg, E.S. (1992). Analysis of early development in the zebrafish embryo. **Results and Problems in Cell Differentiation 18**, 91-150.

Werner, M., and Burley, S. (1997). Architectural transcription factors: proteins that remodel DNA. **Cell 88**, 733-736.

Westerfield, M. (ed.) (1994). **The Zebrafish Book**. University of Oregon Press, Eugene, Oregon.

Wilmut, M., Schnieke, A.E., McWhir, J., kind, A.J., and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. **Nature 385**, 810-813.

Wong, K.C.Y. (1995/1996) Characterisation of a zebrafish (Danio Rerio) bZIP transcription factor belonging to the PAR subfamily. **Thesis for the degree of Bachelor of Science with Honours.** National University of Singapore, Singapore.

Errata

1. The definition of the leucine zipper (p16, 9th line)

Leucine zippers consist of a stretch of amino acids with a leucine residue in every seventh position. The interaction between two proteins containing this domain form an intermolecule coiled-coil dimer that stabilizes (or "zips up") the partners.

2. The nature of the probe used to screen the library (p56, 1st line)

During the course of setting up conditions for cDNA library screening using digoxigenin-labeled DNA probes, we picked up a false-positive clone from a 20-28 hpf embryonic cDNA library using the zebrafish beta-actin probe (a kind gift of Dr Gong ZY).